

STIC-Biotech/ChemLib

97013

**From:** Slobodyansky, Elizabeth  
**Sent:** Thursday, June 19, 2003 12:24 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/889,609

Please search for case 09/889,609:



SEQ ID NOS: 1 and 8-10 against commercial ~~and interference~~ databases.

Thank you.

*Elizabeth Slobodyansky, PhD*

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

**mail box 10D01**

Mary Jane Ruhl  
Tech. Info. Specialist, STIC  
TC-1600  
CM-1, Room 6A-06  
Phone: 605-1155

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: June 24, 2003, 06:24:52 ; Search time 2668 Seconds  
(without alignments)  
11158.984 Million cell updates/sec

Title: US-09-889-609B-9

Perfect score: 1023  
Sequence: 1 atgtgtctcagttattgcat.....atcagtcgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number\_of\_hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_btg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_vi.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1023	100.0	1023	9	AY034081	AY034081 Homo sapi
2	1023	100.0	1166	9	AF169974	AF169974 Homo sapi
3	1023	100.0	1770	6	AX224418	AX224418 Sequence
4	1023	100.0	2477	9	AK023169	AK023169 Homo sapi
5	1021.4	99.8	1336	6	AX259575	AX259575 Sequence
6	1020	99.7	1020	6	AX224420	AX224420 Sequence
7	867	84.8	2264	9	HS0805450	AL834378 Homo sapi
8	816.6	79.8	1197	10	AF148321	AF148321 Mus muscu
9	816.6	79.8	1411	10	BC011164	BC011164 Mus muscu
10	221	21.6	3807	9	BC019090	BC019090 Homo sapi
11	221	21.6	4107	9	AB037822	AB037822 Homo sapi
12	221	21.6	4140	9	AK026565	AK026565 Homo sapi
13	221	21.6	123454	9	HSBC17A99	AL450226 Homo sapi
14	221	21.6	221909	2	AC021705	AC021705 Homo sapi
15	181.6	17.8	184151	10	AL604066	AL604066 Mouse DNA
16	181.6	17.8	205910	2	AC044847	AC044847 Mus muscu
17	180	17.6	3379	10	BC031531	BC031531 Mus muscu
18	176.2	17.2	101814	2	AC122087	AC122087 Rattus no
19	157.6	15.4	40005	8	SPCC320	AL022245 S.pombe c
20	157.6	15.4	42689	8	SPCC330	AL031603 S.pombe c
21	157.2	15.4	6525	8	SCYK1218C	228218 S.cerevisia
22	157.2	15.4	24577	8	SC6ORF	X75951 S.cerevisia
23	142	13.9	295350	1	AP004826	AP004826 Staphyloc
24	137	13.4	137	11	G32814	G32814 A009527 Hum
25	131.4	12.8	298050	1	AP003189	AP003189 Clostridi
26	128.4	12.6	291050	1	AP000982	AP000982 Sulfolobu
27	126.2	12.3	16931	1	AE001716	AE001716 Thermotog
28	126	12.3	303750	1	AP003133	AP003133 Staphyloc
29	126	12.3	346900	1	AP003362	AP003362 Staphyloc
30	114	11.1	348050	1	AP003581	AP003581 Nostoc sp
31	111.4	10.9	303040	1	AP000991	AP000991 Thermopla
32	110.4	10.8	9990	1	AE010646	AE010646 Fusobacte
33	110.2	10.8	157609	2	AC095280	AC095280 Rattus no
34	105	10.3	268200	1	AP000061	AP000061 Aeropyrum
35	105	10.3	317511	1	CJ11168X3	AL139076 Campyloba
36	104	10.2	1631	3	AY052059	AY052059 Drosophil
37	100.2	9.8	35437	1	SCF43A	AL096837 Streptomy
38	98.4	9.6	166050	1	AL646085	AL646085 Ralstonia
39	97.8	9.6	14449	1	AE004696	AE004696 Pseudomon
40	96.6	9.4	1779	8	AF096281	AF096281 Arabidops
41	96.6	9.4	2056	8	AF177212	AF177212 Arabidops
42	94.4	9.2	10805	1	AE011664	AE011664 Xanthomon
43	94.4	9.2	10805	1	AE011664	AE011664 Xanthomon
44	92.8	9.1	8044	8	SPBC1677	AL035581 S.pombe c
45	89.4	8.7	10881	1	U32757	U32757 Haemophilus

ALIGNMENTS

RESULT 1	AY034081	AY034081	1023 bp	linear	PRI 01-JUN-2002
LOCUS	AY034081	Homo sapiens serine racemase mRNA, complete cds.			
DEFINITION	AY034081				
ACCESSION	AY034081				
VERSION	AY034081.1	GI:21307620			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 1023)				
	Xia,M., Liu,Y. and Connolly,T.M.				
	Molecular cloning and expression of human serine racemase from				
	human NT2N'cells				

Unpublished  
 2 (bases 1 to 1023)  
 XIA.M., Liu, Y. and Connolly, T.M.  
 Direct Submission  
 Submitted (08-MAY-2001) Pharmacology, Merck & Co., Summeytown Pike,  
 West Point, PA 19486, USA  
 Location/Qualifiers  
 1..1023  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17p13"  
 1..1023  
 /codon\_start=1  
 /product="serine racemase"  
 /protein\_id="AAK8495.1"  
 /db\_xref="GI:21307621"  
 /translation="MCAQYCSIFADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLF  
 KCELFQKTSFKIRGNALNVRSLVDPALERKPAVTVHSSNGHQAALTYAAKLEGIPA  
 YIVVPTAPDCFKLAIOAYGASIVYCEPDESRENKRVTEETEGIMVHPNQPAVI  
 AGQGTIALEVLNOVPLDALYVPVGGGMLAGIATVKAALKPSKVYVAEPENADDCY  
 OSKLGKLMPLNLPETIADGVKSSIGLNTWPIIRDLDVDDIFTVEDEIKCATOLVWE  
 RMKLLIEPTAGVGAVALSQHFQIVSPVKNICIVLSGNDVLLSSITWVQAERPAS  
 YQSVSV"  
 BASE COUNT 290 a 220 c 243 g 270 t  
 ORIGIN  
 Query Match 100.0%; Score 1023; DB 9; Length 1023;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-266;  
 Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGTGTCAGTATGTCATCTCCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 DB 1 ATGTGTCAGTATGTCATCTCCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 QY 61 GATTCATCCACCTACACCACTGCTACACAGCTCCATTTTGATCACTAACAGGGGC 120  
 DB 61 GATTCATCCACCTACACCACTGCTACACAGCTCCATTTTGATCACTAACAGGGGC 120  
 QY 121 AATCTTTCTTCAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTCT 180  
 DB 121 AATCTTTCTTCAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTCT 180  
 QY 181 CTCATGCGGTGAGAAGCTTGTTTCTCGATGCTTTAGAAAAGAACGCGAAGCTGTGTT 240  
 DB 181 CTCATGCGGTGAGAAGCTTGTTTCTCGATGCTTTAGAAAAGAACGCGAAGCTGTGTT 240  
 QY 241 ACTCACAGCAGTGAACCACTGCGCAGGCTCTCACCTATGCTGCCAAATGGAGGAATT 300  
 DB 241 ACTCACAGCAGTGAACCACTGCGCAGGCTCTCACCTATGCTGCCAAATGGAGGAATT 300  
 QY 301 CCTGCTTATATTGTGGTGCCCGACAGCTCCAGACTGTAAAAAAGCTTGAATACAAGCC 360  
 DB 301 CCTGCTTATATTGTGGTGCCCGACAGCTCCAGACTGTAAAAAAGCTTGAATACAAGCC 360  
 QY 361 TACGAGGCGCAATTGTATATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420  
 DB 361 TACGAGGCGCAATTGTATATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420  
 QY 421 AGAGTTACAGAAAACAGAGGATCATGTATATCCACAGGAGCCTGCAGTGATA 480  
 DB 421 AGAGTTACAGAAAACAGAGGATCATGTATATCCACAGGAGCCTGCAGTGATA 480  
 QY 481 GCTGACAAGGACAATTGGCCCTGGAAGTGTGAACCAAGGTTCTTTGGTGGATGCACTG 540  
 DB 481 GCTGACAAGGACAATTGGCCCTGGAAGTGTGAACCAAGGTTCTTTGGTGGATGCACTG 540  
 QY 541 GTGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600  
 DB 541 GTGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 600  
 QY 601 AAACCTAGTGTAGGATATATGCTGTGAACCTCAATGCAGATGACTCTACCAGTCC 660  
 DB 601 AAACCTAGTGTAGGATATATGCTGTGAACCTCAATGCAGATGACTCTACCAGTCC 660

DB 601 AAACCTAGTGTAGGATATATGCTGTGAACCTCAATGCAGATGACTGTACCAGTCC 660  
 QY 661 AAGCTGAAGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC 720  
 DB 661 AAGCTGAAGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC 720  
 QY 721 AAATCCAGCATTTGGCTTGAACACCTGGGCTATTATCATGGGACCTTTGTGGATGATATCTTC 780  
 DB 721 AAATCCAGCATTTGGCTTGAACACCTGGGCTATTATCATGGGACCTTTGTGGATGATATCTTC 780  
 QY 781 ACTGTACAGAGGATGAATAAAGTGTGCAACCCAGCTGGTGGGAGAGGATGAAACTA 840  
 DB 781 ACTGTACAGAGGATGAATAAAGTGTGCAACCCAGCTGGTGGGAGAGGATGAAACTA 840  
 QY 841 CTCATTGAACCTACAGCTGGTGGTGGAGTGGCTGCTGCTCTCACACATTTTCAAACT 900  
 DB 841 CTCATTGAACCTACAGCTGGTGGTGGAGTGGCTGCTGCTCTCACACATTTTCAAACT 900  
 QY 901 GTTTCCCCAGAAAGTAAAGAACATTTGATTGTGCTGCTGCTGGAATGTAGACTTAACC 960  
 DB 901 GTTTCCCCAGAAAGTAAAGAACATTTGATTGTGCTGCTGCTGGAATGTAGACTTAACC 960  
 QY 961 TCCTCCATAACTTGGTGAACAGGCTGAAAGCCAGCTTCTTATCATGCTGTGTTCTGTT 1020  
 DB 961 TCCTCCATAACTTGGTGAACAGGCTGAAAGCCAGCTTCTTATCATGCTGTGTTCTGTT 1020  
 QY 1021 TAA 1023  
 DB 1021 TAA 1023

RESULT 2  
 AFI69974  
 LOCUS Homo sapiens serine racemase mRNA, complete cds.  
 DEFINITION AFI69974  
 ACCESSION AFI69974.1 GI:11034784  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 De Miranda, J., Santoro, A., Engelen, S. and Wolosker, H.  
 Human serine racemase: molecular cloning, genomic organization and  
 functional analysis  
 Gene 256 (1-2), 183-188 (2000)  
 JOURNAL  
 MEDLINE 20510003  
 PUBMED 11054547  
 REFERENCE 2 (bases 1 to 1166)  
 DeMiranda, J., Santoro, A.H., Engelen, S. and Wolosker, H.  
 Direct Submission  
 Submitted (18-JUL-1999) Departamento de Bioquímica Médica,  
 Instituto de Ciencias Biomédicas, Universidade Federal do Rio de  
 Janeiro, Av. Brigadeiro Trompowski, s./n, Rio de Janeiro 21491-590,  
 Brazil

FEATURES  
 source  
 1..1166  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="brain"  
 17..1039  
 /function="catalyzes L- to D-serine racemization"  
 /codon\_start=1  
 /product="serine racemase"  
 /protein\_id="AAG27081.1"  
 /db\_xref="GI:11034785"  
 /translation="MCAQYCSIFADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLF  
 KCELFQKTSFKIRGNALNVRSLVDPALERKPAVTVHSSNGHQAALTYAAKLEGIPA  
 YIVVPTAPDCFKLAIOAYGASIVYCEPDESRENKRVTEETEGIMVHPNQPAVI  
 AGQGTIALEVLNOVPLDALYVPVGGGMLAGIATVKAALKPSKVYVAEPENADDCY  
 OSKLGKLMPLNLPETIADGVKSSIGLNTWPIIRDLDVDDIFTVEDEIKCATOLVWE  
 RMKLLIEPTAGVGAVALSQHFQIVSPVKNICIVLSGNDVLLSSITWVQAERPAS

CDS



YQSVSV"																							
BASE COUNT	333 a	238 c	275 g	320 t																			
ORIGIN																							
Query Match 100.0%; Score 1023; DB 9; Length 1166;																							
Best Local Similarity 100.0%; Pred. No. 6.5e-266;																							
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;																							
QY	1	ATGTGTCCTCAGTATTGCAATCTCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA	60																				
DB	17	ATGTGTCCTCAGTATTGCAATCTCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA	76																				
QY	61	GATTCTATCCCTCACACAGGCTGCTAAAGCTCCATTTTGAATCAACATAACAGGCGC	120																				
DB	77	GATTCTATCCCTCACACAGGCTGCTAAAGCTCCATTTTGAATCAACATAACAGGCGC	136																				
QY	121	AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTAAGATTCGTGTGCT	180																				
DB	137	AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTAAGATTCGTGTGCT	196																				
QY	181	CTCAATGCCGTGAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCGGAAAGCTGTGTT	240																				
DB	197	CTCAATGCCGTGAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCGGAAAGCTGTGTT	256																				
QY	241	ACTCAGCAGTGGAAACCAATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT	300																				
DB	257	ACTCAGCAGTGGAAACCAATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT	316																				
QY	301	CTGTCTATATCTGTGTCGCCAGACAGCTCCAGACTGTAAAAAAGCTTGAATACAGCC	360																				
DB	317	CTGTCTATATCTGTGTCGCCAGACAGCTCCAGACTGTAAAAAAGCTTGAATACAGCC	376																				
QY	361	TACGGAGCGTCAATTTGATCTGTAACCTAGTGTAGTCCAGAGAAATTTGCAAAA	420																				
DB	377	TACGGAGCGTCAATTTGATCTGTAACCTAGTGTAGTCCAGAGAAATTTGCAAAA	436																				
QY	421	AGAGTTACAGAGAAACAGAGGATCATGGTATACATCCCAACAGGAGCTGCAGTGATA	480																				
DB	437	AGAGTTACAGAGAAACAGAGGATCATGGTATACATCCCAACAGGAGCTGCAGTGATA	496																				
QY	481	GCTGGACAAGGGACAATTCCTGGAAGTCTGAACAGGTTCTCTTTGGTGGATGCACTG	540																				
DB	497	GCTGGACAAGGGACAATTCCTGGAAGTCTGAACAGGTTCTCTTTGGTGGATGCACTG	556																				
QY	541	GTGCTACCTGTAGTGGAGGAGGAATCTTCTGGAATAGCAATACAGTTAAGGCTCTG	600																				
DB	557	GTGCTACCTGTAGTGGAGGAGGAATCTTCTGGAATAGCAATACAGTTAAGGCTCTG	616																				
QY	601	AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAAATGACATGACTGCTACCACTCC	660																				
DB	617	AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAAATGACATGACTGCTACCACTCC	676																				
QY	661	AACTGAAGGGAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGCTGTC	720																				
DB	677	AACTGAAGGGAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGCTGTC	736																				
QY	721	AAATCCAGCATTTGGCTTGAAACCTGGCCCTATTATCAGGAGACCTTGTGGATGATATCTTC	780																				
DB	737	AAATCCAGCATTTGGCTTGAAACCTGGCCCTATTATCAGGAGACCTTGTGGATGATATCTTC	796																				
QY	781	ACTGTACAGAGGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACATA	840																				
DB	797	ACTGTACAGAGGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACATA	856																				
QY	841	CTCATTTGACCTACAGCTGTGTGGAGTGGCTGCTGTGCTGTCTCAACATTTTCAACT	900																				
DB	857	CTCATTTGACCTACAGCTGTGTGGAGTGGCTGCTGTGCTGTCTCAACATTTTCAACT	916																				
QY	901	GTTTTCCCGAGAGTAAGAACATTTTGTATTGTCTCAGTGGTGAATGTAGACTTTAACC	960																				
DB	917	GTTTTCCCGAGAGTAAGAACATTTTGTATTGTCTCAGTGGTGAATGTAGACTTTAACC	976																				
QY	961	TCCTCCATAACTTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTCTCTGTT	1020																				

Db	977	TCCTCCATAA	1023	1770 bp	DNA	linear	PAT 10-SEP-2001
QY	1021	TAA	1039	Sequence 1	from Patent WO0160987		
Db	1037	TAA	1039	AX224418	AX224418	GI:15554660	
RESULT 3	AX224418	LOCUS	AX224418	DEFINITION	Sequence 1	from Patent WO0160987	
DEFINITION	AX224418	VERSION	AX224418.1	GI:15554660			
KEYWORDS		SOURCE	human				
ORGANISM			Homo sapiens				
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE			1 (bases 1 to 1770)				
AUTHORS			Meyers, R.A. and Rudolph-Owen, L.A.				
TITLE			A human pyridoxal-phosphate dependent enzyme family member and uses therefor				
JOURNAL			Patent: WO 0160987-A 1 23-AUG-2001;				
FEATURES			Millennium Pharmaceuticals, Inc. (US)				
source			1. 1770				
			/organism="Homo sapiens"				
			/db_xref="taxon:9606"				
CDS			69..1091				
			/note="unnamed protein product"				
			/codon_start=1				
			/protein_id="CAC69570.1"				
			/db_xref="GI:15554661"				
			/translation="MCAQYGISFADVEKAKHINIRDSIHLTPVLTSILNOLTGRNLFPA				
			KCELFQKTSFKIRNALNVRSLVPDALERKPAKVTHSSGNHGOALTYAAKLEGPVIV				
			VIVPQTAPDCKLAIQAYGASIVCEPSDESRENKRVTEGTEGIMVHNPQEPVIV				
			AQGGTIAILEVNLQVLDALVPVPGSGMLAGIAITVKALKPSVKYVAAPSPNADDCY				
			OSKLGKLMPLNYPETIADGVKSSIGLITWPIIRLDVDDIFVTVEDEIKCATOLVME				
			RMKLLIETPAGVGAAYLVLSQHFQTVSPVKNKICIVLSGNGVNDLTSSITWVQKAERPA				
			YQSVSV"				
BASE COUNT	499 a	379 c	410 g	482 t			
ORIGIN							
Query Match	100.0%;	Score 1023;	DB 6;	Length 1770;			
Best Local Similarity	100.0%;	Pred. No. 6.4e-266;					
Matches 1023;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
QY	1	ATGTGTCCTCAGTATTGCAATCTCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA	60				
DB	69	ATGTGTCCTCAGTATTGCAATCTCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCGA	128				
QY	61	GATTCTATCCCTCACACAGGCTGCTTACCAAGCTCCATTTTGAATCAACATAACAGGCGC	120				
DB	129	GATTCTATCCCTCACACAGGCTGCTTACCAAGCTCCATTTTGAATCAACATAACAGGCGC	188				
QY	121	AATCTTTTCTCAAAATGTGAATCTTCCAGAAAAACAGGATCTTTTAAAGATTCGTGTGCT	180				
DB	189	AATCTTTTCTCAAAATGTGAATCTTCCAGAAAAACAGGATCTTTTAAAGATTCGTGTGCT	248				
QY	181	CTCAATGCCGTGAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCCGAAAGCTGTGTTGTT	240				
DB	249	CTCAATGCCGTGAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCCGAAAGCTGTGTTGTT	308				
QY	241	ACTCAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTT	300				
DB	309	ACTCAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTT	368				
QY	301	CCTGCTTATATTGTGTCGCCCGACAGACTCTCCAGACTGTAAAAAAGCTTGAATACAGCC	360				
DB	369	CCTGCTTATATTGTGTCGCCCGACAGACTCTCCAGACTGTAAAAAAGCTTGAATACAGCC	428				
QY	361	TACGGAGGCTCAATTTGTATTACTGTGAACCTAGTGTAGTCCAGAGAAAAATGTTGCCAAA	420				

Db	977	TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT	1036																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
----	-----	--	------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

|||||
429 TACGGAGCGTCAATGTGTATATCTGTGAACCTAGTGTGATGAGTCCAGAGAAATGTTGCAAAA 488
QY 421 AGAGTTACAGAGAAACAGAGGCGCATGTGTATCATCCCAACAGGAGCGCTGCAGTGATA 480
Db 489 AGAGTTACAGAGAAACAGAGGCGCATGTGTATCATCCCAACAGGAGCGCTGCAGTGATA 548
QY 481 GCTGGACAGGAGCAATTTGCCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 540
Db 549 GCTGGACAGGAGCAATTTGCCCTGGAAGTGTGTAACAGGTTCTTTGGTGGATGCACTG 608
QY 541 GTGGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTGTG 600
Db 609 GTGGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTGTG 668
QY 601 AAACCTAGTGTGAAGGTATATGCTGTAACCTCAATCAGATGACTGCTACCACTGC 660
Db 669 AAACCTAGTGTGAAGGTATATGCTGTAACCTCAATCAGATGACTGCTACCACTGC 728
QY 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 720
Db 729 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 788
QY 721 AAATCCAGCATGCTTGAACACTGGCTTATATCAGGAGACCTTGTGGATGATATCTTC 780
Db 789 AAATCCAGCATGCTTGAACACTGGCTTATATCAGGAGACCTTGTGGATGATATCTTC 848
QY 781 ACTGTCTACAGAGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTA 840
Db 849 ACTGTCTACAGAGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTA 908
QY 841 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTGCTGCTCAACATTTTCAAACT 900
Db 909 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTGCTGCTCAACATTTTCAAACT 968
QY 901 GTTTCCTCCAGCAAGTAAAGACATTTGATTGTCTCAGTGGTGAATGTAGACTTAACC 960
Db 969 GTTTCCTCCAGCAAGTAAAGACATTTGATTGTCTCAGTGGTGAATGTAGACTTAACC 1028
QY 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1020
Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGTT 1088
QY 1021 TAA 1023
Db 1089 TAA 1091

RESULT 4
AK023169
LOCUS
DEFINITION
Homo sapiens CDNA FLJ13107 fls, clone NT2RP3002501, weakly similar
to THREONINE DEHYDRATASE CATABOLIC (EC 4.2.1.16).
ACCESSION
AK023169
VERSION
1 GI:10434971
KEYWORDS
oligo capping; fls (full insert sequence).
SOURCE
Homo sapiens
clone lib: NT2RP3 clone: NT2RP3002501.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
TITLE
Homo sapiens cDNA NT2RP3002501
REFERENCE
2 (bases 1 to 2477)
AUTHORS
Isogai,T. and Otsuki,T.

```

```

Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3002501"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
CDS
69..1091
/notes="unnamed protein product"
/codon_start=1
/protein_id="BABI4442.1"
/db_xref="GI:10434972"
/translation="MCAQYICSPADVEKAHINIRDSIHLTPVLTSSILNLTGRNLEFF
KSELFORTSFKIRGALNAVRSILVDALEKRPRKAVKRVTSNHNHQAITYAAKLEGIPA
YIVVPTAPDCKKLAIQAGASTVYCEPSEDSRENKAKRVTEETEGIVAAAPSPADDCY
AGGCTALEVLNQVPLVDALVVPVGGMLAGIATVKALKPKSVKVAAPSPADDCY
QSKLKKLMPNLYPPETIADGVKSSIGLNTPIRLVDIDFTVTEDEIKCATQLVNE
RMKLLLEPTAGVGAVALSOHFQTVSPVKNICIVLSSGNDVLTSSITWVQAEFPAS
YQSVSV"
BASE COUNT 738 a 535 c 524 g 680 t
ORIGIN
Query Match 100.0%; Score 1023; DB 9; Length 2477;
Best Local Similarity 100.0%; Pred. No. 6.3e-266;
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTGTGCTCAGTATTGCAATCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCGA 60
Db 69 ATGTGTGCTCAGTATTGCAATCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCGA 128
QY 61 GATTCTATCCACCTCACACAGTCTTCAACAGCTCCATTTTGAATCAACTAACAGGCGC 120
Db 129 GATTCTATCCACCTCACACAGTCTTCAACAGCTCCATTTTGAATCAACTAACAGGCGC 188
QY 121 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 180
Db 189 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGT 248
QY 181 CTCGAATGCCGTGAGAAGCTTGGTTCCTGTGATGCTTTTGAAGAAAGCGGAAAGCTGTGTT 240
Db 249 CTCGAATGCCGTGAGAAGCTTGGTTCCTGTGATGCTTTTGAAGAAAGCGGAAAGCTGTGTT 308
QY 241 ACTCACAGCAGTGGAAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300
Db 309 ACTCACAGCAGTGGAAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 368
QY 301 CTGCTTTATATTGTTGGTGGCCCAACAGAGCTTCCAGAGCTGTAAAAAAGCTTCAATACAGGCC 360
Db 369 CTGCTTTATATTGTTGGTGGCCCAACAGAGCTTCCAGAGCTGTAAAAAAGCTTCAATACAGGCC 428
QY 361 TAGCGAGCGTCAATTTGATCTGTAACCTAGTGTAGTGTGAGTCCAGAGAAAATGTTGCAAAA 420
Db 429 TAGCGAGCGTCAATTTGATCTGTAACCTAGTGTAGTGTGAGTCCAGAGAAAATGTTGCAAAA 488
QY 421 AGAGTTTACAGAGAAACAGAGGCGCATGTGTTACATCCCAACAGGAGCGCTGCAGTGATA 480
Db 489 AGAGTTTACAGAGAAACAGAGGCGCATGTGTTACATCCCAACAGGAGCGCTGCAGTGATA 548

```

QY	481	GCTGACAGGAGCAATTCCTGGAAGTGGTGAACAGGTTCCCTTTGGTGGATGCACTG	540
Db	549	GCTGACAGGAGCAATTCCTGGAAGTGGTGAACAGGTTCCCTTTGGTGGATGCACTG	608
QY	541	GTGGTACCTGTAGGTGGAGGAGGAATGCTGCTGGAATAGCAATTAAGTAAAGGCTCTG	600
Db	609	GTGGTACCTGTAGGTGGAGGAGGAATGCTGCTGGAATAGCAATTAAGTAAAGGCTCTG	668
QY	601	AAACCTAGTGTGAAGGTATATGCTGCAACCCCTCAATGCAGATGACTGTACCACTCC	660
Db	669	AAACCTAGTGTGAAGGTATATGCTGCAACCCCTCAATGCAGATGACTGTACCACTCC	728
QY	661	AAGTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG	720
Db	729	AAGTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG	788
QY	721	AAATCCAGCATTCGCTTGAACACCTGGCTTATTATCAGGAGCCTTGTGGATGATCTTC	780
Db	789	AAATCCAGCATTCGCTTGAACACCTGGCTTATTATCAGGAGCCTTGTGGATGATCTTC	848
QY	781	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA	840
Db	849	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA	908
QY	841	CTCAATTGAACCTACAGCTGGTGTGGAGTGCTGCTGTGCTCAACATTTTCAAACT	900
Db	909	CTCAATTGAACCTACAGCTGGTGTGGAGTGCTGCTGTGCTCAACATTTTCAAACT	968
QY	901	GTTCCTCCAGAGTAAGAAATTTGATTTGCTGCTAGTGGTGAATGTAGACTTAACT	960
Db	969	GTTCCTCCAGAGTAAGAAATTTGATTTGCTGCTAGTGGTGAATGTAGACTTAACT	1028
QY	961	TCCTCCATAACTTGGGTGAACAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1020
Db	1029	TCCTCCATAACTTGGGTGAACAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1088
QY	1021	TAA 1023	
Db	1089	TAA 1091	
RESULT 5			
AX259575			
LOCUS	AX259575	1336 bp	DNA linear PAT 26-OCT-2001
DEFINITION	Sequence 1 from Patent WO0173077.		
ACCESSION	AX259575		
VERSION	AX259575.1 GI:16508702		
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 Ramakrishnan,S.		
JOURNAL	Regulation of human serine racemase enzyme		
FEATURES	Patent: WO 0173077-A 1 04-OCT-2001;		
source	Bayer Aktiengesellschaft (DE)		
BASE COUNT	Location/Qualifiers		
ORIGIN	1. 1336		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	386 a	268 c	319 g 362 t 1 others
Query Match 99.8%; Score 1021.4; DB 6; Length 1336;			
Best Local Similarity 99.9%; Pred. No. 1.8e-265;			
Matches 1022; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	ATGTGTGCTCAGTATTGCTATCTCTTCTGCTGATGTTGAAAGAGCTCATATCAACATTCGA	60
Db	52	ATGTGTGCTCAGTATTGCTATCTCTTCTGCTGATGTTGAAAGAGCTCATATCAACATTCGA	111
QY	61	GATTCTATCCACCTCACACCACTGCTACAAAGCTCCATTTTGAATCAACTAACAGGGCGC	120

Db	112	GATTCTATCCACCTCACACCACTGCTAAACAGCTCCATTTTGAATCAACTAACAGGGCGC	171
QY	121	AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGCT	180
Db	172	AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAACAGGATCTTTTAAGATTCGTGGTGCT	231
QY	181	CTCAATGCCGTGCAAGCTTGGTCTCTGATGCTTTAGAAAAGGAGCCGAAAGCTGTTCGT	240
Db	232	CTCAATGCCGTGCAAGCTTGGTCTCTGATGCTTTAGAAAAGGAGCCGAAAGCTGTTCGT	291
QY	241	ACTCACAGCAGTGAAGAACCAATGGCCAGGCTCTACCTATATGCTGCCAAATTTGGAAGGAAT	300
Db	292	ACTCACAGCAGTGAAGAACCAATGGCCAGGCTCTACCTATATGCTGCCAAATTTGGAAGGAAT	351
QY	301	CCTGCTTATATTTGGTGGCCCCCAGACACTCCAGACTGTAAAAAACTTTGCAATACAAGCC	360
Db	352	CCTGCTTATATTTGGTGGCCCCCAGACACTCCAGACTGTAAAAAACTTTGCAATACAAGCC	411
QY	361	TACGAGGCGTCAATTTGATAGTGAACCTAGTATGATGATGATGATGATGATGATGATGATG	420
Db	412	TACGAGGCGTCAATTTGATAGTGAACCTAGTATGATGATGATGATGATGATGATGATGATG	471
QY	421	AGAGTTACAGAAAGAAACAGAGGATCATGTTGATATCCCAACAGGAGCCCTGCAGTGATA	480
Db	472	AGAGTTACAGAAAGAAACAGAGGATCATGTTGATATCCCAACAGGAGCCCTGCAGTGATA	531
QY	481	GCTGCAGAGGAGCAATTTGCCCTGGAAGTGTGATGATGATGATGATGATGATGATGATGATG	540
Db	532	GCTGCAGAGGAGCAATTTGCCCTGGAAGTGTGATGATGATGATGATGATGATGATGATGATG	591
QY	541	GTGCTACCTGTAGGTGGAGGAGGAATGCTGCTGGAATAGCAATTTACAGTTTAAGGCTCTG	600
Db	592	GTGCTACCTGTAGGTGGAGGAGGAATGCTGCTGGAATAGCAATTTACAGTTTAAGGCTCTG	651
QY	601	AAACCTAGTGTGAAGGTATATGCTGCTGGAACCCCTCAAAATGCAGATGACTGTACCACTCC	660
Db	652	AAACCTAGTGTGAAGGTATATGCTGCTGGAACCCCTCAAAATGCAGATGACTGTACCACTCC	711
QY	661	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC	720
Db	712	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC	771
QY	721	AAATCCAGCATTGGCTTGAACACCTGGCTATTATCAGGAGCCTTTGTGGATGATATCTTC	780
Db	772	AAATCCAGCATTGGCTTGAACACCTGGCTATTATCAGGAGCCTTTGTGGATGATATCTTC	831
QY	781	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA	840
Db	832	ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGGATGAAACTA	891
QY	841	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	900
Db	892	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	951
QY	901	GTTCCTCCAGAGTAAGAAATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	960
Db	952	GTTCCTCCAGAGTAAGAAATTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1011
QY	961	TCCTCCATAACTTGGGTGAACAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1020
Db	1012	TCCTCCATAACTTGGGTGAACAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT	1071
QY	1021	TAA 1023	
Db	1072	TAA 1074	
RESULT 6			
AX224420			
LOCUS	AX224420	1020 bp	DNA linear PAT 11-SEP-2001
DEFINITION	Sequence 3 from Patent WO0160987.		
ACCESSION	AX224420		

```

VERSION      AX224420.1  GI:15554662
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE        1 (bases 1 to 1020)
             A human pyridoxal-phosphate dependent enzyme family member and uses
             therefor
JOURNAL      Patent: WO 0160987-A 3 23-AUG-2001;
             Millennium Pharmaceuticals, Inc. (US)
FEATURES     Location/Qualifiers
             1..1020
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT   288 a 220 c 243 g 269 t
ORIGIN
Query Match      99.7%; Score 1020; DB 6; Length 1020;
Best Local Similarity 100.0%; Pred. No. 4.3e-265;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTGCTCAGTATGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTTCGA 60
Db 1 ATGTGCTCAGTATGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTTCGA 60
QY 61 GATTCATATCCACCTCACACAGTGTCTAACAAGCTCCATTGTAATCAACTACAGGGCGC 120
Db 61 GATTCATATCCACCTCACACAGTGTCTAACAAGCTCCATTGTAATCAACTACAGGGCGC 120
QY 121 AATCTTTTCTTCAAATGTGAATCTTCCAGAAAACAGGATCTTTTAAGATTCTGCTGCT 180
Db 121 AATCTTTTCTTCAAATGTGAATCTTCCAGAAAACAGGATCTTTTAAGATTCTGCTGCT 180
QY 181 CTCATATGCGCTCAGAAGCTTGTTCCCTGATGCTTTAGAAAGAACGCCGAAAGCTTGT 240
Db 181 CTCATATGCGCTCAGAAGCTTGTTCCCTGATGCTTTAGAAAGAACGCCGAAAGCTTGT 240
QY 241 ACTCACAGCAGTGGAACCATGCGCCAGCTCTACCTATGCTGCCAAATTTGAAGGAATT 300
Db 241 ACTCACAGCAGTGGAACCATGCGCCAGCTCTACCTATGCTGCCAAATTTGAAGGAATT 300
QY 301 CTGCTTATATGTGGTCCCGCCAGACAGCTCCAGACTGTAAAAAACCTTGCATACAAGCC 360
Db 301 CTGCTTATATGTGGTCCCGCCAGACAGCTCCAGACTGTAAAAAACCTTGCATACAAGCC 360
QY 361 TACGAGGCGTCAATTTGATATCTGTAACCTAGTAGTATGATCCAGAGAAAATGTTGCAAAA 420
Db 361 TACGAGGCGTCAATTTGATATCTGTAACCTAGTAGTATGATCCAGAGAAAATGTTGCAAAA 420
QY 421 AGAGTTACAGAGAAGAACAGAGGCATCATGGTATACATCCCAACAGGCGCTGCAGTGATA 480
Db 421 AGAGTTACAGAGAAGAACAGAGGCATCATGGTATACATCCCAACAGGCGCTGCAGTGATA 480
QY 481 GCTGACAAAGGACAAATTCCTGGAAGTGTGAACACAGGTTCTCTTTGGTGGATGCACTG 540
Db 481 GCTGACAAAGGACAAATTCCTGGAAGTGTGAACACAGGTTCTCTTTGGTGGATGCACTG 540
QY 541 GTGGTACCTGTAGGTGGAGGAGGAGTCTGCTGGAATAGCAATACAGTTAAAGCTCTG 600
Db 541 GTGGTACCTGTAGGTGGAGGAGGAGTCTGCTGGAATAGCAATACAGTTAAAGCTCTG 600
QY 601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATATGCAGATGACTGTACCAAGTCC 660
Db 601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATATGCAGATGACTGTACCAAGTCC 660
QY 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATACAGATGGTGTG 720
Db 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATACAGATGGTGTG 720
QY 721 AAATCCAGCATTGGCTTCAACACCTGGCTATTATCAGGACCTTGTGGATGATATCTTC 780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```

Db 721 AAATCCAGCATTGGCTTGAACACCTGGCTATTATCAGGAGCCTTGTGGATGATATCTTC 780
QY 781 ACTGTCACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGGGAGAGATGAACATA 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 ACTGTCACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGGGAGAGATGAACATA 840
QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
QY 901 GTTTCCCCAGAGTAAGAACATTTGTATTTGTCTAGTGGTGGGAAATGTAGACTTAAC 960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 GTTTCCCCAGAGTAAGAACATTTGTATTTGTCTAGTGGTGGGAAATGTAGACTTAAC 960
QY 961 TCCTCCATAAATCTGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTCTTTCTGTT 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 TCCTCCATAAATCTGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTCTTTCTGTT 1020

RESULT 7
HSM805450      2264 bp  mRNA  linear  PRI 12-JUL-2002
LOCUS          Homo sapiens mRNA; cDNA DKFZp762A2415 (from clone DKFZp762A2415).
DEFINITION     AL834378
ACCESSION      AL834378.1  GI:21740048
VERSION        AL834378.1
KEYWORDS       human.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 2264)
AUTHORS        Ansoerge,W., Wirkner,U., Mewes,H.W., Weill,B. and Wiemann,S.
TITLE          Direct Submission
JOURNAL        Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT        Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
               Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
               Heidelberg/Germany) within the cDNA sequencing consortium of the
               German Genome Project.
               This clone (DKFZp762A2415) is available at the RZPD in Berlin.
               Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
               Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
               information about the clone and the sequencing project is available
               at http://mips.gsf.de/proj/cDNA/.

FEATURES       Location/Qualifiers
               1..2264
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /map="22.9 cR from top of Chr17 linkage group"
               /clone="DKFZp762A2415"
               /tissue_type="melanoma (MeWo cell line)"
               /clone_lib="762 (synonym: hmel2). Vector pSport1; host
               DH10B; sites NotI + SalI"
               /dev_stage="adult"
               1..2264
               /gene="DKFZp762A2415"
               <1..867
               /note="similarity to serine racemase (Mus musculus)"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="CAD39041.1"
               /db_xref="GI:21740049"
               /translation="GSPKIRGALNAVRLSLVDALEKPKAVVTHSSGNHGOALTVAAK
               LEGIPAVIVPQTAPDCKKLAIOAGSIVCEPDESREINAKRVETEGIMVHPN
               QEPAVIAGQGTIALEVLNQVELDALVYPVGGGMLAGIATVAKLKFSKVVYAAEPS
               NADCYOSKLGKLEMPNLYPPETIADGVKSSIGLNTWPIIRLDVDDIETVDEIKCA
               TOLYWERKLLIEPTAGVGAVALSQQHFQTVSPVKNICIIVSGNVDLTSSITWVKQ
               AERPASYQSVSV"
               2226..2231
               /polyA_signal
               2253
               /polyA_site
               /gene="DKFZp762A2415"
```

AF148321.1		GI:6448864	
VERSION	Mus musculus.		
KEYWORDS	Mus musculus		
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
ORGANISM	Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1197)		
AUTHORS	Wolosker, H., Blackshaw, S. and Snyder, S.H.		
TITLE	Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13409-13414 (1999)		
MEDLINE	20027561		
PUBMED	10557334		
REFERENCE	2 (bases 1 to 1197)		
AUTHORS	Wolosker, H.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-MAY-1999) Neuroscience, Johns Hopkins University, 725 North Wolfe St., WBSB 806, Baltimore, MD 21205, USA		
FEATURES	Location/Qualifiers		
source	1..1197		
	/organism="Mus musculus"		
	/strain="BALB/c"		
	/db_xref="taxon:10090"		
CDS	66..1085		
	/EC_number="5.1.1.17"		
	/function="catalyzes L-serine to D-serine racemization"		
	/codon_start=1		
	/product="serine racemase"		
	/protein_id="AAF08701.1"		
	/db_xref="GI:6448865"		
	/translation="MCAQYCI SFADVEKAHINIQDSIHLTPVLTSILNQIAGRLFFKCELFQKSGFKIRGALNAIRGLIPDTPEEKPAVTVHSSNGHQAATVAAKLEGPAYIVVPTAPNCKLQATQVASTIVCDPDESEKVTORIMQETGILVHPNQEPVIAQGGTALVNLQVPLVDALVYPVGGGHWAGIATIKALKPSYKVAEAPSADDCYQSKLGEPLNLHPPTIADGVKSSIGLNTWPIIRDLDVDDFTVTDEDEIKYATQLVGKMKLIEPTAGVALAAVLQSHQFTVSPKVNIVLSSGNGVLDLSLNVWVQAEARPAFYKTVSV"		
BASE COUNT	345 a 283 c 283 g 286 t		
ORIGIN			
Query Match	79.8%; Score 816.6; DB 10; Length 1197;		
Best Local Similarity	88.1%; Pred. No. 5e-210;		
Matches	901; Conservative 0; Mismatches 119; Indels 3; Gaps 1;		
QY	1 ATGTGTGCTCAGTATTGCTATCCCTTTGCTGTGATGTTGAAAAAGCTCATATCAACATTCGA 60		
Db	66 ATGTGTGCTCAGTACTGCTCTCTCTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 125		
QY	61 GATTCATCCACTCACACAGCTGCTTAACAAGCTCCATTTTGAATCAACTAACAGGCGCG 120		
Db	126 GACTCTATCCACTCACCCAGTGTCTAACAAGCTCCATTTTGAATCAATAGCAGGCGCG 185		
QY	121 AATCTTTTCTCAATGTGAACCTTCCAGAAACAGGATCTTTTAAAGATTCGAGTGGCT 180		
Db	186 AATCTTTTCTCAATGTGAGCTCTTCCAGAAAACTGGGTCCTTTTAAAGATTCGAGTGGC 245		
QY	181 CTCAATGCGCTCAGAGCTTGGTTCCTGATGCTTTTGAAGAGGAGCGAAGCTGTGTTGTT 240		
Db	246 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCGCCAGAGAGAGCCCAAGCGCTAGTT 305		
QY	241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATT 300		
Db	306 ACTCACAGCAGCGAAACCATGGCCAGGCTCTCACCTATGCTGCTAAACTGGAAGGAATT 365		
QY	301 CCGTCTTATTTGTTGGTCCCGACAGAGCTCCAGACTGTAAAAAAGCTTGCATACAGACC 360		
Db	366 CCGTCTTACATTTGTTGGTTCCCGAACACAGCTCCCAACTGCCAAGAACTGGCAATCCAAGC 425		
QY	361 TACGGAGCGCTCAATTTGATATCTGTAACCTAGTAGTATGATGATGATGATGATGATGAT 420		
Db	426 TATGGAGCATCGATAGTATATCTGTGACCCAAAGTACAGAGTCCAGAGAAAGGTCACCTCAA 485		
QY	421 AGAGTTACAGAGAAACAGAGGATCATGTTGATATCTCCCAACAGGAGGCTGTCAGTGATA 480		



Db 768 GCTGGACAGCAAAATGCTCCCTGGAGTCTCTGAACAGGTTCCCTTGGTAGATGCACTG 827  
 QY 541 GTGCTACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAACAGTTAAAGGCTCTG 600  
 Db 828 GTGGTACCAGTACGAGGAGGAGGAATGCTTGTGGAATAGCAATTAACAGTTAAAGGCTCTG 887  
 QY 601 AAACCTAGTGTAGGATATATGCTGTGAACCTCAAAATCAGATGACTCTACCACTCC 660  
 Db 888 AAACCTAGTGTAGGATATATGCTGTGAACCTCAAAATCAGATGACTCTACCACTCT 947  
 QY 661 AAGCTGAAGGGGAAATGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGCTGTC 720  
 Db 948 AAATGAAAGAGAGACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGCTGTC 1007  
 QY 721 AAATCCAGCATTTGGCTTGAACACTGCGCTTATATCAGGAGCACTTGTGGATGATATCTTC 780  
 Db 1008 AAATCCAGCATTTGGCTTGAATACCTGCGCTTATATCAGGAGCACTTGTGGATGATCTTC 1067  
 QY 781 ACTGTCCACAGAGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACACTA 840  
 Db 1068 ACTGTCCACAGAGATGAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACACTG 1127  
 QY 841 CTCATTGAACCTACAGCTGTGTGGAGTGGCTGTGTCTGCTCAACATTTTCAAACT 900  
 Db 1128 CTCATTGAGCCGACTGCTGCGTGGCACTGGCTGCACTGTCTCAGCATTTTCCAAACA 1187  
 QY 901 GTTTCCTCCAGAGTAAAGAACTTTGTTATTTGCTGCTCAGTGTGGAAATGTAGACTTAACC 960  
 Db 1188 GTCTCTCCAGAGTAAAGAACTTTGTTATTTGCTGCTCAGTGTGGAAATGTAGACTTAA 1245  
 QY 961 TCCTCCATACCTTGGGTGAAGAGCTGGAAGCCAGCTCTTATCATCAGTCTGTTCTGTT 1020  
 Db 1246 -CTCCTCGTAAGTGGGTGGGCGAGGTGAACGCCAGCTCTTACCAGAGGTTTCTGTT 1304  
 QY 1021 TAA 1023  
 Db 1305 TAA 1307

RESULT 10  
 BC019090/c  
 LOCUS 3807 bp mRNA linear PRI 11-DEC-2001  
 DEFINITION Homo sapiens, clone MGC:29514 IMAGE:4896567, mRNA, complete cds.  
 ACCESSION BC019090  
 VERSION BC019090.1 GI:17512217  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Direct Submission  
 Submitted (07-DEC-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: [http://www.nisc.nih.gov/nisc\\_mgc/nhgri.nih.gov](http://www.nisc.nih.gov/nisc_mgc/nhgri.nih.gov)  
 Contact: Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamen, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: [http://www.nisc.nih.gov/nisc\\_mgc/nhgri.nih.gov](http://www.nisc.nih.gov/nisc_mgc/nhgri.nih.gov)  
 Contact: Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamen, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tlionson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 39 Row: 3 Column: 10  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 8922495.

FEATURES  
source

Location/Qualifiers  
 1..3807

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:29514 IMAGE:4896567"  
 /tissue\_type="Brain, neuroblastoma"  
 /clone\_lib="NIH\_MGC\_19"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"

## CDS

319..2289  
 /codon\_start=1  
 /product="unknown (protein for MGC:29514)"  
 /protein\_id="AAH19090.1"  
 /db\_xref="GI:17512218"  
 /translation="MAKVADTILFLDPLEGWDSTGDYCLSLCLFAOGLPTVTLAVOGI  
 SGLPKKQIDFRKKLSKAVEKREPHDKLLLDLTDQOERALLROLANGKQKOLAFDRRR  
 AYLFHAFVDFPSENNLVGLKISGYVRGOTLNNRLLHIVGDFQOMKQIDAPGDP  
 FPLNPRGIRPKQKPDMAEICATDAVDMEELKVLKMDPGRQESLQAEVIFPDMEG  
 EOTWPEELSEAKDFLEGSVKVVKPKGTSSVQAEWILDGSEAKMLEYKOELEEE  
 FMDPEVDTPDVAARIRFOYKRLGKSFRTSPMDPKENLPDYARIFQFONFTNRKSI  
 FKEVEKEVEGAEGVTVLHVSEVPVVECFRQGTPLIAFSLLPHEQKMSVLNMVY  
 RRDPGNTEPVKAEELIFHCGFRFRASPLFSQHTAADKHLQRFUTADMALVATYA  
 PITFPASVLLFKQKNGMHSLLATGHLMSVDPDPMVIRKVLVSGHPFKITRKAVVR  
 YMFNREDVLWFKPVELRTKWRGRGHKEPLGTGHGKMSFCGDKLKSQDTVLNLYKR  
 VFPKWTYDVPVPEPVLKSEISSTVPOQGM"

BASE COUNT  
ORIGIN

1102 a 813 c 943 g 949 t

Query Match 21.6%; Score 221; DB 9; Length 3807;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-49;  
 Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 AGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 862  
 Db 2832 AGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 2773  
 QY 863 TTGGAGTGGCTGTGTCTCTCAACATTTTCAAACTGTTTCCCCAGAGTAAGAACA 922  
 Db 2772 TTGGAGTGGCTGTGTCTCTCAACATTTTCAAACTGTTTCCCCAGAGTAAGAACA 2713  
 QY 923 TTTGTATTGTGCTCAGTGGTGGAAATAGAGCTTAACCTCCTCCATTAACCTGGTGAAGC 982  
 Db 2712 TTTGTATTGTGCTCAGTGGTGGAAATAGAGCTTAACCTCCTCCATTAACCTGGTGAAGC 2653  
 QY 983 AGGCTGAAGCCAGCTTCTTATCAGTCTGTTCTGTTTAA 1023  
 Db 2652 AGGCTGAAGCCAGCTTCTTATCAGTCTGTTCTGTTTAA 2612

RESULT 11  
AB037822/c

LOCUS 4107 bp mRNA linear PRI 14-MAR-2000  
 DEFINITION Homo sapiens mRNA for KIAA1401 protein, partial cds.  
 ACCESSION AB037822  
 VERSION AB037822.1 GI:7243182  
 KEYWORDS

## SOURCE

Homo sapiens brain cDNA to mRNA, clone\_lib:pBluescriptII SK plus  
 Clone: hK09639.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Nagase, T., Kikuno, R., Ishikawa, K.I., Hirose, M. and Ohara, O.  
 Prediction of the coding sequences of unidentified human genes.



XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro

DNA Res. 7 (1), 65-73 (2000)

20181126

2 (bases 1 to 4107)

REFERENCE  
Ohara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

JOURNAL  
MEDLINE  
AUTHORS  
TITLE  
JOURNAL

## FEATURES

Location/Qualifiers

Source

1..4107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hk09639"  
/tissue\_type="brain"  
/clone\_lib="pBluescriptII SK plus"

gene

1..4107

CDS

/gene="KIAA1401"  
/note="KIAA1401"  
/codon\_start=1  
/product="KIAA1401 protein"  
/protein\_id="BAA92639.1"  
/db\_xref="GI:7243183"  
/translation="HGALTSCKWRSGAGSPAPWRARPLGDKRLRSLWPSGADSV  
HARGMAARPGPLKQNKAKHGGRHGRGSAORDGKRLALKLTSKVKRKLRSVQ  
RRASOLRKKQKEAVLAELKQGLGKDPGPPHVLVPLHSRSLPEAKQLQDRGTG  
HLNELGTQNPMLCPLKRLKRFETSRAPGDLHVLDMAKVDATILFLDPLEGWDST  
GDYCLSLFAOGLTYLAVOGISLPLKKQIDTRKLSKAVEKRFPHDLKLLDGLTQ  
EAGMLQLANLQOHLAFRRAYLFAHAFVDFPSNNLVGLTKISGVVGGTQV  
NRLHIVYGDSLOAQIDAPDPPLPGIKPKQDPDAMEICATDAVDMDERGLAV  
LMKADPQROSLQAVIPDPMEGQTPWTEELSEAKDFLKESSKVKYKPKGTSSSQ  
AEWLDGSGSGSGEGVEYDDMEHEDFMEESODESESEEEETMTIGESVHDDLY  
DKVDEAEAKMLKYLKQRLKLEFPEVDTPDVAARIRFKYRGLKSPVVECFRQ  
ENLPDARIFQFONFTNTRKLEFPEVEEVEGEAGVYVTLHVSVPVSVVECFRQ  
GTLPLAFSLPHQKMSVLNVRVRDNPTEPPASVLIFKOKSNMHSILFIATGLMSVDPDR  
AADKKLQRLFTADMLAVATVYAPITPPASVLIFKOKSNMHSILFIATGLMSVDPDR  
MVTKRVYLSGHPKFIETKMAVRYMEFNREDVLWFKPVELRTKQWGRGHLKPELTHG  
HMKSPDGKLSQDPTLNLNLYKRVFPEKWTYDYPVPEPVLKSEISSTVPGQME"

BASE COUNT

1094 a 992 c 1131 g 890 t

ORIGIN

Query Match 21.68; Score 221; DB 9; Length 4107;  
Best Local Similarity 100.0%; Pred. No. 8.3e-49;  
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 862  
|||||  
Db 4037 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGT 922  
|||||

QY 863 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCAAGAGTAAGAACA 922  
|||||  
Db 3977 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCAAGAGTAAGAACA 3918  
|||||

QY 923 TTTGTATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 982  
|||||  
Db 3917 TTTGTATTGCTCAGTGGTGGAAATGTAGACTTAACCTCCTCCATAAATTTGGGTGAAGC 3858  
|||||

QY 983 AGGCTGAAGGCCAGCTTCTTATCAGTCTGTTCTGTTTAA 1023  
|||||

Db 3857 AGGCTGAAGGCCAGCTTCTTATCAGTCTGTTCTGTTTAA 3817  
|||||

RESULT 12

AK026565/c

LOCUS

AK026565

DEFINITION

Homo sapiens cDNA: FLJ22912 fis, clone KAT06132.

ACCESSION

AK026565

VERSION

AK026565.1 GI:10439447

4140 bp mRNA linear

Homo sapiens cDNA: FLJ22912 fis, clone KAT06132.

AK026565

AK026565.1 GI:10439447

AK026565

Homo sapiens cDNA: FLJ22912 fis, clone KAT06132.

AK026565

AK026565.1 GI:10439447

## KEYWORDS

SOURCE

ORGANISM

Homo sapiens

## REFERENCE

AUTHORS

Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project

NEDO human cDNA sequencing project



```

REFERENCE
AUTHORS Ramser,J., Langer,I., Klein,M., Arndt,S., Lehrack,S., Junker,E.,
Schuelzchen,S., Noitz,M., Thompson,C., Hoff,C., Poustka,A.,
Reinhardt,R. and Lehrach,H.
JOURNAL
REFERENCE 2 (bases 1 to 123454)
AUTHORS MOLGENR.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) MOLGENR, Abt. Lehrach, Max Planck Institut
Fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany
COMMENT contig 01
FEATURES
source 1. .123454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q13.3"
/clone="PAC RPCI-5 1037N22"
/clone_lib="RPCI1,3-5, Roswell Park Cancer Institute,
creator Pieter de Jong, P.Ioannou"
/notes="region between markers D17S695 - D17S654"
misc_feature 1
/notes="T7_end:PAC RPCI-5 1037N22"
misc_feature 123454
/notes="SP6_end:PAC RPCI-5 1037N22"
BASE COUNT 31233 a 26833 c 28651 g 36737 t
ORIGIN
Query Match 21.6%; Score 221; DB 9; Length 123454;
Best Local Similarity 100.0%; Pred. No. 7e-49; Indels 0; Gaps 0;
Matches 221; Conservative 0; Mismatches 0;
803 AGTGTCAACCCAGCTGGTGGGAGGATGAACCTACTTGAACCTACAGCTGGTG 862
|||||
15555 AGTGTCAACCCAGCTGGTGGGAGGATGAACCTACTTGAACCTACAGCTGGTG 15496
|||||
863 TTGAGTGGCTGCTGCTGCTCAACATTTTCAAACTGTTCCCGAGAGTAAGAACA 922
|||||
15495 TTGAGTGGCTGCTGCTGCTCAACATTTTCAAACTGTTCCCGAGAGTAAGAACA 15436
|||||
923 TTGTATTGTGCTCACTGGTGGAAATGTAGACTTAACCTCCCTCATTAACCTGGTGAAGC 982
|||||
15435 TTGTATTGTGCTCACTGGTGGAAATGTAGACTTAACCTCCCTCATTAACCTGGTGAAGC 15376
|||||
983 AGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 1023
|||||
15375 AGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 15335
|||||

RESULT 14
AC021705/c
LOCUS 221909 bp DNA linear HTG. 23-AUG-2002
DEFINITION Homo sapiens chromosome 17 clone RP11-380H7 map 17, *** SEQUENCING
IN PROGRESS ***, 5 unordered pieces.
AC021705
ACCESSION AC021705.19 GI:22450645
VERSION HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 221909)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-380H7
Unpublished
2 (bases 1 to 221909)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bida,F.,
Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepl,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

```

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Leoczky,J., Levine,R., Lieu,C., Liu,C., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 221909)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepl,Y., Collymore,A., S.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Minova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 23, 2002 this sequence version replaced gi:22296935.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5879
Center clone name: 380_H_7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 49592: contig of 49592 bp in length
* 49593 49692: gap of 100 bp
* 49693 65686: contig of 15994 bp in length
* 65687 65786: gap of 100 bp
* 65787 69248: contig of 3462 bp in length
* 69249 69348: gap of 100 bp
* 69349 92592: contig of 23244 bp in length
* 92593 92692: gap of 100 bp
* 92693 221909: contig of 129217 bp in length.
Location/Qualifiers
1. .221909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-380H7"
/clone_lib="RPCI-11 Human Male BAC"
FEATURES
source

```

BASE COUNT 53146 a 52419 c 54384 g 61545 t 415 others  
ORIGIN

Query Match 21.6%; Score 221; DB 2; Length 221909;  
Best Local Similarity 100.0%; Pred. No. 6.8e-49;  
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 803 AGTGTGCAACCCAGCTGCTGTGGGAGAGGATGAACCTACTCATTTGAACCTACAGCTGGTG 862  
|||||  
Db 74677 AGTGTGCAACCCAGCTGCTGTGGGAGAGGATGAACCTACTCATTTGAACCTACAGCTGGTG 74618  
|||||  
Qy 863 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCCAGAGTAAAGAACA 922  
|||||  
Db 74617 TTGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTTCCCCAGAGTAAAGAACA 74558  
|||||  
Qy 923 TTTCTATTGTGCTCAGTGGTGGGAATGTAGACTTAACCTCTCCATAACTTGGGTGAAGC 982  
|||||  
Db 74557 TTTGATTGTGCTCAGTGGTGGGAATGTAGACTTAACCTCTCCATAACTTGGGTGAAGC 74498  
|||||  
Qy 983 AGGCTGAAGGCCAGCTCTTATCAGTCTGTTCTGTTTAA 1023  
|||||  
Db 74497 AGGCTGAAGGCCAGCTCTTATCAGTCTGTTCTGTTTAA 74457  
|||||

RESULT 15  
AL604066/c  
LOCUS 184151 bp DNA linear ROD 14-MAR-2002  
DEFINITION Mouse DNA sequence from clone Rp23-174M12 on chromosome 11,  
complete sequence.

ACCESSION AL604066  
VERSION AL604066.9 GI:19571943  
KEYWORDS HTG.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (Bases 1 to 184151)  
Bray-Allen, S.  
Direct Submission

Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
On Mar 21, 2002 this sequence version replaced gi:17902958.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
from the RPI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6

This sequence is the entire insert of clone Rp23-174M12 The true  
right end of clone Rp23-194P5 is at 82810 in this sequence.

FEATURES  
source  
Location/Qualifiers  
1..184151  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="Rp23-174M12"

BASE COUNT 46879 a 41019 c 43482 g 52771 t  
ORIGIN  
Query Match 17.8%; Score 181.6; DB 10; Length 184151;  
Best Local Similarity 91.0%; Pred. No. 3.2e-38;  
Matches 193; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Qy 593 AGGCTCTGAACACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAAAATGCAGATGACTGCT 652  
|||||  
Db 125770 AGGCCCTGAACACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAAAATGCAGATGACTGCT 125711  
|||||  
Qy 653 ACCAGTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 712  
|||||  
Db 125710 ACCAGTCTAAACTGAAGGAGAACTGACCCCCCAATCTTCTCCTCCAGAAACCATAGCAG 125651  
|||||  
Qy 713 ATGCTGTCAAACTCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGGACCTTGTGGATG 772  
|||||  
Db 125650 ATGCTGTCAAACTCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGAGACCTTGTGGATG 125591  
|||||  
Qy 773 ATATCTTCACTGTCACAGAGGATGAATTAAG 804  
|||||  
Db 125590 ATGTCTTCACTGTCACCGGAAGATGAATCAAG 125559  
|||||

Search completed: June 24, 2003, 10:29:39  
Job time : 2674 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 04:05:10 ; Search time 1462 Seconds  
(without alignments)  
11277.022 Million cell updates/sec

Title: US-09-889-609B-1

Perfect score: 1018

Sequence: 1 atgtgtctcagctactgcatt.....ttaccagacgtctgttttaa 1018

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154056 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpi:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	996	97.8	1398	11 AK017235	Mus muscu
2	642.4	63.1	769	12 BF163794	BF163794 601769701
3	603.8	59.3	785	13 BF163753	BF163753 603047780
4	599.6	58.9	848	9 AUI131397	AUI131397 AUI131397
5	595.8	58.5	667	10 BB609829	BB609829 BB609829
6	583.8	57.3	600	13 BI988879	BI988879 4022-49 M

7	554.2	54.4	565	10	BE55084
8	543.8	53.4	819	13	BI739030
9	496.2	48.7	918	14	BQ885621
10	479.8	47.1	823	13	BI738966
11	475.2	46.7	698	9	AV022510
12	461.2	45.3	602	10	AV374491
13	460.4	45.2	634	14	BM179814
14	455	44.7	455	9	AI322578
15	445.6	43.8	729	13	BG965678
16	429.2	42.2	793	12	BG404240
17	428.4	42.1	527	9	AA446793
18	418.6	41.1	827	13	BI332919
19	406.4	39.9	556	10	BB692486
20	400.8	39.4	830	12	BF135412
21	393.2	38.6	519	10	BB641631
22	393.2	38.6	652	10	AV709023
23	386.8	38.0	697	10	AV649313
24	367.2	36.1	688	10	BB625125
25	367	36.1	380	9	AA034539
26	366.2	36.0	661	10	BB621410
27	353.8	34.8	1087	12	BF532321
28	334.4	32.8	645	10	AW211242
29	333.8	32.8	646	10	BB644584
30	325.4	32.0	391	14	BQ303565
31	323.8	31.8	1159	14	BQ215405
32	323	31.7	404	14	BQ303566
33	322	31.6	386	14	BQ303566
34	320.2	31.5	657	10	AW319255
35	317.2	31.2	611	10	BB618702
36	316	31.0	632	10	BB641468
37	314.2	30.9	672	10	BE380996
38	300.8	29.5	1045	12	BF384769
39	284	27.9	471	14	H73097
40	283.4	27.8	394	10	AV649514
41	269.8	26.5	742	9	AI525507
42	267.6	26.3	438	14	H86748
43	265.6	26.1	572	10	BB651611
44	253	24.9	698	10	AW550801
45	253	24.9	698	13	BM213835

## ALIGNMENTS

RESULT 1	AK017235	AK017235	1398 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330405D10:serine racemase, full insert sequence.					
DEFINITION	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330405D10:serine racemase, full insert sequence.					
ACCESSION	AK017235					
VERSION	AK017235.1 GI:12856372					
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:5330405D10.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
AUTHORS	1 Carninci, P. and Hayashizaki, Y.					
TITLE	High-efficiency full-length cDNA cloning					
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)					
MEDLINE	99279253					
PUBMED	10349636					
REFERENCE	2					
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.					
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes					
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)					
MEDLINE	20499374					
PUBMED	11042159					

2526/88/1

May



Db 890 AAACCTAGTGTGAAGATATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 949

QY 661 AAACCTGAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGC 720

Db 950 AAACCTGAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGC 1009

QY 721 AAATCCAGATTGGCTTGAATACCTGGCCCTATTATAAGAGAGACTTGTGGATGATGCTTC 780

Db 1010 AAATCCAGATTGGCTTGAATACCTGGCCCTATTATAAGAGAGACTTGTGGATGATGCTTC 1069

QY 781 ACTGTCCAGAGATGAATCAATCAAGTATGAACCCAGCTGGTGGGGGAGAAATGAACATG 840

Db 1070 ACTGTCCAGAGATGAATCAATCAAGTATGAACCCAGCTGGTGGGGGAGAAATGAACATG 1129

QY 841 CTCATTGAGCCGACTGCTGGCTGGCAGCTGGCTGAGTCTGCTCAGCAGATTTCACAAACA 900

Db 1130 CTCATTGAGCCGACTGCTGGCTGGCAGCTGGCTGAGTCTGCTCAGCAGATTTCACAAACA 1189

QY 901 GTCTCTCCAGAGTAAAGAACCTGTGCATTGTACTGAGTGGGGGGAATGTAGACCTAAC 960

Db 1190 GTCTCTCCAGAGTAAAGAACCTGTGCATTGTACTGAGTGGGGGGAATGTAGACCTAAC 1249

QY 961 TCCCTGAACTGGTGGGGGAGCTGAACGGCCAGCTCCTTACCAGACGCTCTGTTT 1016

Db 1250 TCCCTGAACTGGTGGGGGAGCTGAACGGCCAGCTCCTTACCAGACGCTCTGTTT 1305

RESULT 2  
BF163794  
LOCUS 601769701F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:398873 5',  
DEFINITION mRNA sequence.  
ACCESSION BF163794  
VERSION BF163794.1 GI:11043971  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9198 row: a column: 02  
High quality sequence stop: 699.

FEATURES  
source  
1. .769  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C2EHC II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:398873"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
Stem cell origin."  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 225 a 186 c 187 g 171 t

Query Match 63.1%; Score 642.4; DB 12; Length 769;  
Best Local Similarity 98.7%; Pred. No. 3e-183;

Matches 679; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

QY 1 ATGTGTGCTCAGTACTGATCTCCCTTTGCTGTGTTGTTGAAAAAGCTCATATCAACATTCAA 60

Db 40 ATGTGTGCTCAGTACTGATCTCCCTTTGCTGTGTTGTTGAAAAAGCTCATATCAACATTCAA 99

QY 61 GACTCTATCAGCTCACCCAGTGTAAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120

Db 100 GACTCTATCAGCTCACCCAGTGTAAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 159

QY 121 AATCTTTTCTTCAATGTGAGCTCTCCAGAAAACCTGGGTCTTTTAAAGATTCAGGTGCC 180

Db 160 AATCTTTTCTTCAATGTGAGCTCTCCAGAAAACCTGGGTCTTTTAAAGATTCAGGTGCC 219

QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCC - AGAAGAGAAGCCCAAAACCCCTAGT 239

Db 220 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCCCAAGAGAAGAGCCCAAAACCCCTAGT 279

QY 240 TACTCACAGCAGCGGAAACCATGGCCAAAGCTCTCACTATGTCTTAAACTTGAAGTGAAGAAAT 299

Db 280 TACTCACAGCAGCGGAAACCATGGCCAAAGCTCTCACTATGTCTTAAACTTGAAGTGAAGAAAT 339

QY 300 TCTCTGCTTACATTTGGTTTCCCAACACAGCTCCCAACTGCAAGAAACTGGCAATCCAAGC 359

Db 340 TCTCTGCTTACATTTGGTTTCCCAACACAGCTCCCAACTGCAAGAAACTGGCAATCCAAGC 399

QY 360 CTATGGAGCATCGATAGTATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGTCACTCA 419

Db 400 CTATGGAGCATCGATAGTATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGTCACTCA 459

QY 420 AAGAATTATGCAAGAAACAGAGGCACTTTGGTTCCTCCCAACAGAGGCGCTGCAGTGTAT 479

Db 460 AAGAATTATGCAAGAAACAGAGGCACTTTGGTTCCTCCCAACAGAGGCGCTGCAGTGTAT 519

QY 480 AGCTGGACAAGAACAAATTTGCCCTGGAAGTGTGACCAAGGTTCCCTTGGTAGATGCAT 539

Db 520 AGCTGGACAAGAACAAATTTGCCCTGGAAGTGTGACCAAGGTTCCCTTGGTAGATGCAT 579

QY 540 GGTGGTACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTAACATTAAGGCCCT 599

Db 580 GGTGGTACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTAACATTAAGGCCCT 638

QY 600 GAAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGTCAGATGCTGTACCAGTGC 659

Db 639 GAAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTTGAATGC - GATGACTGGTACCAGTC 697

QY 660 TAAACTGAAGAGAGAACTGACCCCAAT 687

Db 698 TAAACTGAAGAGAGAACTGACCCCAAT 725

RESULT 3  
BF163753  
LOCUS 603047780F1 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:518099 5',  
DEFINITION mRNA sequence.  
ACCESSION BF163753  
VERSION BF163753.1 GI:15755331  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 785)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11469 row: p column: 20

High quality sequence stop: 783.

Location/Qualifiers

#### FEATURES

source  
1. .785  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5188099"  
/clone\_lib="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dr primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

BASE COUNT  
ORIGIN

225 a 168 c 200 g 192 t  
Query Match 59.3%; Score 603.8; DB 13; Length 785;  
Best Local Similarity 87.9%; Pred. No. 1.5e-171;  
Matches 602; Conservative 0; Mismatches 92; Indels 3; Gaps 3;  
QY 104 ATCAATAGCAGGGCCCAATCTTTCTTCAATGTGAGCTCTTCCAGAACTGGGTCTT 163  
Db 1 ATCAACTAAGCAGGGCCCAATCTTTCTTCAATGTGAGCTCTTCCAGAAACAGGATCTT 60  
QY 164 -TTAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATTCCTGACACGCCAGAGAG 222  
Db 61 GTTAAGATTCGAGGTGCCCTTAATGCCCTCAGAGGCTTGGTTCCTGATCTTTAGAAAGG 120  
QY 223 AAGCCCAAGCCGTAGTTACTCACAGCAGGAAACCATGGCAAGCTCTCACTATGCT 282  
Db 121 AAGCCGAAGCTGT-GTTACTCACAGCATGGAACCATGGCCAGGCTCTCACCATGCT 179  
QY 283 GCTAACTGGAAGGAATTCCTGCTTACATTTGGTTTCCCAACAGCTCCCAACTGCAAG 342  
Db 180 GCCAATTTGGAAGGAATTCCTGCTTATATTGGTGGCCGCCAGACAGCTCCAGACTGTA 239  
QY 343 AACTGGCAATCAAGCCCTTAGGACCATCGATAGTATATCTGTGACCAAGTCAAGATCC 402  
Db 240 AACTTTGCAATCAAGCCCTTAGGAGCGGTCAATTTGTATATCTGTGAACCTAGTATG 299  
QY 403 AGAGAAAGGTCACCTCAAGAAATTTATGCAAGAAACAGAGGATCTTTGCTGATCCCAAC 462  
Db 300 AGAGAAATGTTGCAAAAGAGTTACAGAGAAACAGAGGATCATGTTATATCCCAAC 359  
QY 463 CAGGAGCCCTGAGTATAGCTGGACAGGAACAATTTGCCCTGGAAGTGTGTAACCAAGT 522  
Db 360 CAGGAGCCCTGAGTATAGCTGGACAGGAACAATTTGCCCTGGAAGTGTGTAACCAAGT 419  
QY 523 CCTTTGGTAGTGCAGTGGTGTACCATGAGGAGGAGGAATGGTTCCTGGAATAGCC 582  
Db 420 CCTTTGGTAGTGCAGTGGTGTACCTGTGAGTGGAGGAGGAATGGTTCCTGGAATAGCA 479  
QY 583 ATTACAAATTAAGCCCTGAACCTTAGTGTGAAGGTATACGGCTGTGAGCCCTCGAATGCA 642  
Db 480 ATTACAGTTAAGGCTCTGAAACCTTAGTGTGAAGGTATATGCTGTGAACCCCTCAATGCA 539  
QY 643 GATGACTGCTACAGTCTAAACTGAAAGAGAGACTGACCCCAATCTTTCATCCTCCAGAA 702  
Db 540 GATGACTGCTACAGTCTCAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCCTCCAGAA 599  
QY 703 ACCATAGCAGATGGTGTCAATTCAGCATTGCTTGAATACCTGGCTATTATATAGACAC 762  
Db 600 ACCATAGCAGATGGTGTCAATTCAGCATTTGGCTTGAACACCTGGCTATTATATAGGAC 659

QY 763 CTTGTGATGATGCTCTTCACTGTCACCGAAGATGAATCAAGTATCAACCCAGCTGGTG 822  
Db 660 CTTGTGATGATGATCTTCACTGTCACAGAGATGAATTAAGTGTGC-ACCCAGCTGGTG 718  
QY 823 TGGGGGAGAAATGAACCTGCTCAATTGAGCGGAGCTGCTGGGGTGGCACTGGCTGCAGTCTG 882  
Db 719 TGGGAGAGGATGAACCTACTCATTTGAACCTACAGCTGGTGTGAGAGTGGCTGCTGTGCTG 778  
QY 883 TCTCAGC 889  
Db 779 TCTCAAC 785  
RESULT 4  
LOCUS AUI31397 NT2RP3 Homo sapiens cDNA clone NT2RP3002501 5', mRNA  
DEFINITION AUI31397  
ACCESSION AUI31397  
VERSION AUI31397.1 GI:10991751  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 848)  
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.  
TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
FEATURES  
Location/Qualifiers  
1. 848  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002501"  
/clone\_lib="NT2RP3"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"  
BASE COUNT 236 a 187 c 214 g 207 t  
ORIGIN 4 others  
Query Match 58.9%; Score 599.6; DB 9; Length 848;  
Best Local Similarity 87.6%; Pred. No. 3e-170;  
Matches 687; Conservative 0; Mismatches 93; Indels 4; Gaps 3;  
QY 1 ATGTGTGCTCAGTACTGTCATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
Db 69 ATGTGTGCTCAGTATTGTCATCTCTTGTGATGTTGAAAAAGCTCATATCAACATTCAA 128  
QY 61 GATCTATCCACCTCACCCAGTGTGTACCAAGTCCATTTTGAATCAATAGCAGGCCG 120  
Db 129 GATTCTATCCACCTCACCCAGTGTGTACCAAGTCCATTTTGAATCAATAGCAGGCCG 188  
QY 121 ATCTTTCTTCAATATGTGAGCTCTTCCAGAAAACCTGGTCTTTTAAGATTTCAGGTGCC 180  
Db 189 AATCTTTCTTCAATATGTGAGCTCTTCCAGAAAACAGGATCTTTTAAGATTTCAGGTGCT 248  
QY 181 CTTAATGCCATCAGAGGCTTAATTCTCTGACGCCCAAGAGAGCCCGTAGTT 240

```

Db      249 CTCATGCGCGTCAGAGCTTGTTCTCTGATGCTTTAGAAAGGAGCCGAAGCTGTGTT 308
QY      241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTGTAAACTGCAAGGAATT 300
Db      309 ACTCAGCAGGTGGAAACCATGCGCAGGCTCTCACCTATGCTGTCCAAATTTGGAAGAAAT 368
QY      301 COTGCTTACATTTGTTGTTCCCAACAGCTCCCAACTGCAAGAAACTGGCAATCCAAAGCC 360
Db      369 CCGTCTTATTTGTTGGTCCCGCCAGACAGCTCCAGAGCTGTAAAAAACCTTGCAATACAAGCC 428
QY      361 FATGAGCATCATAGTAGTACTGTGACCCCAAGTACAGAGTCCAGAGAAAAGTCTACTCAA 420
Db      429 TACGGAGCGTCAATTGTATCTGTGAACCTAGTGTAGTCCAGAGAAAATTTGCAAAA 488
QY      421 AGAATTTATGCAAGAAACAGAGGCTCTTGTGTCATCCCAACCCAGGAGCTCCAGTGATA 480
Db      489 AGATTTACAGAAACAGAGGCTCATGTTGATATCCCAACCCAGGAGCTCCAGTGATA 548
QY      481 GCTGACAAGGAAACAAATTTGCCCTGGAAGTGTCTGAACCGAGTTCCCTTTGGTAGATGCACGTG 540
Db      549 GCTGACAAGGAAACAAATTTGCCCTGGAAGTGTCTGAACCGAGTTCCCTTTGGTGTGATGCACGTG 608
QY      541 GTGGTACCATGATGAGGAGGAGGAATGTTGCTGGAATAGCCATACAAATTAAGGCCCTTG 600
Db      609 GTGGTACCTGTAGTGGGAAAGAAATGCTTGTGGAATAGCAATTAACAGTTAAGGCTCTG 668
QY      601 AAACCTAGTGTGAAGTATACCTGCTGAGCCCTCGAATGTCAGATGCTGCTACCAAGTCT 660
Db      669 AAACCTAGTGTGAAGTATATCTGCTGAACCTCAATGTCAGATGCTGCTGCTGCTGCT 727
QY      661 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db      728 AAGCTGAAGGAGAACTGATGCCCAATCTTTATCTCTCAGAA-CCATACAGATGGTGTG 785
QY      721 AAATCAGCATTTGGCTGGAATACCTGGCTTATTAAGAGACCTTTGTTGGATGATGCTTTC 780
Db      786 AAATCCA-CATTGGCTTGAACACCTGGCTGCTTATATCANGGACCTTGGNGATGATATCTTT 844
QY      781 ACTG 784
Db      845 ACTG 848

```

## RESULT 5

BB609829

LOCUS

BB609829 BB609829 RIKEN full-length enriched, 18 days embryo Mus musculus  
cDNA clone 1190028F16 5', mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.  
Location/Qualifiers  
1. .667  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="1190028F16"  
/clone\_lib="RIKEN full-length enriched, 18 days embryo"  
/dev\_stage="18 days embryo"  
/lab\_host="SOLR"  
/note="Site\_1: XhoI; Site\_2: SstI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCGCAACTCGAGTGTGTTTATTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SstI."  
BASE COUNT 190 a 163 c 166 g 147 t 1 others  
ORIGIN  
Query Match 58.5%; Score 595.8; DB 10; Length 667;  
Best Local Similarity 99.5%; Pred. No. 3.7e-169;  
Matches 597; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTGTCCTCAGTACTGCTATCTCTTGTGCTGATGTTGAAAAGCTCATATCAACATCAA 60  
Db 68 ATGTGTCCTCAGTACTGCTATCTCTTGTGCTGATGTTGAAAAGCTCATATCAACATCAA 127  
QY 61 GACTCTATCCACCTCACCCTGCTGCTTAAACAGCTCCATTTTGAATCAATAGCAGGCGC 120  
Db 128 GACTCTATCCACCTCACCCTGCTGCTTAAACAGCTCCATTTTGAATCAATAGCAGGCGC 187  
QY 121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACCTGCTTTTAAAGATTCAGGTGCC 180  
Db 188 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACCTGCTTTTAAAGATTCAGGTGCC 247  
QY 181 CTTAATGCCATCAGAGGCTTAATCTTCCAGCCAGAGAGAGCCCAAGCCGTAGTT 240  
Db 248 CTTAATGCCATCAGAGGCTTAATCTTCCAGCCAGAGAGAGCCCAAGCCGTAGTT 307  
QY 241 ACTCACAGCAGCGGAAACCATGGCCCAAGCTCTCACCTATGCTGCTAAAGGAATTT 300



```

Db      308 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGGAAGAAAT 367
QY      301 CCTGCTTACATTTCTGTTCCCAACAGCTCCCAACTGCAAGAACTGCAATCCCAAGCC 360
Db      368 CCTGCTTACATTTCTGTTCCCAACAGCTTACCCTGCAAGAACTGCAATCCCAAGCC 427
QY      361 TATGGAGCATCGATAGTACTATCTGTGACCAAGTGGAGAGTCCAGAGAAAAGGTCACTCAA 420
Db      428 TATGGAGCATCGATAGTACTATCTGTGACCAAGTGGAGAGTCCAGAGAAAAGGTCACTCAA 487
QY      421 AGAATTATCGAAGAACAGAGCATCTTGGTCCATCCCAACAGAGGCTCGCAGTGATA 480
Db      488 AGAATTATCGAAGAACAGAGCATCTTGGTCCATCCCAACAGAGGCTCGCAGTGATA 547
QY      481 GCTGGACAAGGAAACATTTGCCCTGGAAAGTCTGCAACCAAGTTCCTTGGTAGATGCACATG 540
Db      548 GCTGGACAAGGAAACATTTGCCCTGGAAAGTCTGCAACCAAGTTCCTTGGTAGATGCACATG 607
QY      541 GTGGTACCAAGTGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAAGGCCCTG 600
Db      608 GTGGTACCAAGTGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAAGGCCCTG 667

```

```

RESULT 6
BI988879
LOCUS      4022-49 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BI988879
VERSION    1
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 600)
JOURNAL    Mu, X., Zhao, S., Perishad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
MEDLINE    White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
COMMENT    Gene expression in the developing mouse retina by EST sequencing
           and microarray analysis
           Nucleic Acids Res. 29 (24), 4983-4993 (2001)
           21671825
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

```

```

FEATURES
source
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
Note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps.
(Manniatidis); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTCAATCTGAGTG---. Other
information regarding entire library may be found at
http://pga.swmed.edu/data/Libraries/microarray_cdna_library.htm."
BASE COUNT 177 a 145 c 134 g 136 t 8 others
ORIGIN

```

```

Query Match 57.3%; Score 583.8; DB 13; Length 600;
Best Local Similarity 98.3%; Pred. No. 1.5e-165;
Matches 585; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

QY      1 ATGTGTGCTCAGTACTGCTATCCCTTCTGCTGATGTTGAAAGAGTCAATCAACATTCAA 60

```

```

Db      6 ATGTGTGCTCAGTACTGCTATCCCTTCTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 65
QY      61 GACTCTATCCACTCACCCAGTCTGCTAACAAGCTCATTGTAATCAATAGCAGGCGC 120
Db      66 GACTNTATCACCTCACCCAGTCTGCTAACAAGCTCATTGTAATCAANAGCAGGCGC 125
QY      121 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCCAGGTGCC 180
Db      126 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCCAGGTGCC 185
QY      181 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCCAGAGAGCCCAAGCCGTAGTT 240
Db      186 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCCAGAGAGCCCAAGCCGTAGTT 245
QY      241 ACTCAGCAGCGGAAACCATGGCCAAAGCTCTCACCTATGCTGCTAACTGGAGGAATTT 300
Db      246 ACTCAGCAGCGGAAACCATGGCCAAAGCTCTCACCTATGCTGCTAACTGGAGGAATTT 305
QY      301 CTGCTTTACATTTGTTTCCCAACAGCTCCCAACTGCAAGAAACCTGGCAATCCCAAGCC 360
Db      306 CTGCTTTACATTTGTTTCCCAACAGCTCCCAACTGCAAGAAACCTGGCAATCCCAAGCC 365
QY      361 TATGGAGCATCGATAGTACTATCTGTGACCAAGTGGAGAGTCCAGAGAAAAGGTCACTCAA 420
Db      366 TATGGAGCATCGATAGTACTATCTGTGACCAAGTGGAGAGTCCAGAGAAAAGGTCACTCAA 425
QY      421 AGAATTATCGAAGAACAGAGCATCTTGGTCCATCCCAACAGAGGCTCGCAGTGATA 480
Db      426 AGAATTATCGAAGAACAGAGCATCTTGGTCCATCCCAACAGAGGCTCGCAGTGATA 485
QY      481 GCTGGACAAGGAAACATTTGCCCTGGAAAGTGTGCAACAGGTTCCCTTGGTAGATGCACATG 540
Db      486 GCTGGACAAGGAAACATTTGCCCTGGAAAGTGTGCAACAGGTTCCCTTGGTAGATGCACATG 545
QY      541 GTGTGTACCAAGTGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAGG 595
Db      546 GTGTGTNCCANTAGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAGG 600

```

```

RESULT 7
BI655084
LOCUS      565 bp mRNA linear EST 06-SEP-2000
DEFINITION UI-M-ALI-ahr-f-12-0-UI.r2 NIH_BMAP_MCO_N Mus musculus cDNA clone
ACCESSION BE655084
VERSION    1
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 565)
JOURNAL    Bonaldo, M.F., Lennon, G. and Soares, M.B.
MEDLINE    Normalization and subtraction: two approaches to facilitate gene
COMMENT     discovery
           Genome Res. 6 (9), 791-806 (1996)
           9704477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1..565

```



```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AL1-ahr-f-12-0-UI"
/clone_lib="NIH_BMAP_MCO_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MCO_N library is a normalized library constructed
from mouse cortex. The tag is a string of 5 nucleotides
present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Ms. Annie Novakovich, Zivic-Miller
Laboratories."
BASE COUNT      152 a 132 c 152 g 129 t
ORIGIN

Query Match      54.4%; Score 554.2; DB 10; Length 565;
Best Local Similarity 99.5%; Pred. No. 1.4e-156;
Matches 556; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 458 CCAACCAGGAGCTGCAGTGTAGCTGGACAGGACAAATTCCTCGGAAGTCTGAACC 517
DB 1 CCAACCAGGAGCTGCAGTGTAGCTGGACAGGACAAATTCCTCGGAAGTCTGAACC 60
QY 518 AGGTTCCCTTGTAGATGCACGTGGTACCAGTAGGAGGAGGAGGAATGGTGTCTGGA 577
DB 61 AGGTTCCCTTGTAGATGCACGTGGTACCAGTAGGAGGAGGAGGAATGGTGTCTGGA 120
QY 578 TAGCCATTACAATTAAAGGCCCTGAAACCTAGTGTGAAGTATACGCTGCTGAGCCCTCGA 637
DB 121 TAGCCATTACAATTAAAGGCCCTGAAACCTAGTGTGAAGTATACGCTGCTGAGCCCTCGA 180
QY 638 ATGCAGATGACGTCTACCACTTAACCTAAGAGGAGAACTGACCCCAATCTTCATCTC 597
DB 181 ATGCAGATGACGTCTACCACTTAACCTAAGAGGAGAACTGACCCCAATCTTCATCTC 240
QY 698 CAGAAACCATAGCAGATGGTGCAATCCAGCATTGGCTTGAATACCTGGCCCTATTATAA 757
DB 241 CAGAAACCATAGCAGATGGTGCAATCCAGCATTGGCTTGAATACCTGGCCCTATTATAA 300
QY 758 GAGACCTTGTGATGATGCTTTCACTGTCCCGAAGATGAATCAAGTATGAACCCAGC 817
DB 301 GAGACCTTGTGATGATGCTTTCACTGTCCCGAAGATGAATCAAGTATGAACCCAGC 360
QY 818 TGGTGTGGGGAGATGAACCTGCTCATTGACCCGACTGCTGGCGTGGCAGTGGCTGCAG 877
DB 361 TGGTGTGGGGAGATGAACCTGCTCATTGACCCGACTGCTGGCGTGGCAGTGGCTGCAG 420
QY 878 TGCTGTCTCAGCATTTCCAAACAGTCTCTCCAGAAAGTAAAGAACGCTCTCATTTGTACTCA 937
DB 421 TGCTGTCTCAGCATTTCCAAACAGTCTCTCCAGAAAGTAAAGAACGCTCTCATTTGTACTCA 480
QY 938 GTGGGGGGAATGTAGACCTTAACCTCCCTGAACCTGGGTGGGGGAGGCTGAACGCCAGCTC 997
DB 481 GTGGGGGGAATGTAGACCTTAACCTCCCTGAACCTGGGTGGGGGAGGCTGAACGCCAGCTC 540
QY 998 CTTACACAGACGCTCTGTTT 1016
DB 541 CTTACACAGACGCTCTGTTT 559

RESULT 8
BI739030
LOCUS
DEFINITION
BI739030
ACCESSION
BI739030
VERSION
BI739030.1
KEYWORDS
EST.
SOURCE
house mouse.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: The Cepko Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11933 row: p column: 19
High quality sequence stop: 817.

FEATURES
Location/Qualifiers
1..819
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5367042"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      236 a 208 c 192 g 183 t
ORIGIN

Query Match      53.4%; Score 543.8; DB 13; Length 819;
Best Local Similarity 99.6%; Pred. No. 2.4e-153;
Matches 545; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTGTGCTCAGTACTGCATCTCTTGTGTGTTGAAAAAGCTCATATCAACATTCAA 60
DB 273 ATGTGTGCTCAGTACTGCATCTCTTGTGTGTTGAAAAAGCTCATATCAACATTCAA 332
QY 61 GACTCTATCCACTCACCCTCAGTGTACAAAGCTCCATTTTGAATCAATAGCAGGCGC 120
DB 333 GACTCTATCCACTCACCCTCAGTGTACAAAGCTCCATTTTGAATCAATAGCAGGCGC 392
QY 121 AATCTTTTCTTCAATGTGAGCTCTCCAGAAACTGGGCTCTTTTGAAGATTCCAGGTGCC 180
DB 393 AATCTTTTCTTCAATGTGAGCTCTCCAGAAACTGGGCTCTTTTGAAGATTCCAGGTGCC 452
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 240
DB 453 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGCCCAAGCCGTAGTT 512
QY 241 ACTACAGAGCGGGAACCATGGCCAGCTCTCACCTTACCTTGTCTGCTAACTGGAAGGAAT 300
DB 513 ACTACAGAGCGGGAACCATGGCCAGCTCTCACCTTACCTTGTCTGCTAACTGGAAGGAAT 572
QY 301 CCGTCTTACATTGTGGTCTCCCAACAGCTCCCACTGCAAGAACTGCAATCCAGCC 360
DB 573 CCGTCTTACATTGTGGTCTCCCAACAGCTCCCACTGCAAGAACTGCAATCCAGCC 632
QY 361 TATGGAGCATCGATAGTATATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 420
DB 633 TATGGAGCATCGATAGTATATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 692
QY 421 AGAATTATCGAAGAAACAGAGGCAATCTTGGTCCATCCCAACAGAGAGCTCCAGTGATA 480
DB 693 AGAATTATCGAAGAAACAGAGGCAATCTTGGTCCATCCCAACAGAGAGCTCCAGTGATA 752
QY 481 GCTGGACAGGAAACAAATTCCTTGGAGTGTGTAACCAAGGTTCCCTTGGTAGATGCACATG 540
DB 753 GCTGGACAGGAAACAAATTCCTTGGAGTGTGTAACCAAGGTTCCCTTGGTAGATGCACATG 812

```

```

- Qy      541 GTGGTAC 547
      |||||
Db      813 GTGGTAC 819

RESULT 9
BQ885621
LOCUS      918 bp      mRNA      linear      EST 16-AUG-2002
DEFINITION      AGENCOURT_8754012 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332599
5', mRNA sequence.
ACCESSION      BQ885621
VERSION      BQ885621.1 GI:22277639
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
      house mouse.

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 918)
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-remail.nih.gov
      Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
      Ph.D.
      cDNA Library Preparation: ResGen, Invitrogen Corp
      DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Agencourt Bioscience Corporation
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM13790 row: h column: 08
      High quality sequence stop: 609.

FEATURES
      Location/Qualifiers
          1..918
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="IMAGE:6332599"
              /clone_lib="NIH_MGC_130"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: otcysts; Vector: pCMV-SPORT6.1.cdbb;
              Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
              Primer: Oligo dT. Average insert size 1.95 kb.
              Constructed by ResGen, Invitrogen Corp. Note: this is a
              NIH_MGC Library."
          247 a      230 c      230 g      203 t      8 others
BASE COUNT      247 a      230 c      230 g      203 t      8 others
ORIGIN
Query Match      48.7%; Score 496.2; DB 14; Length 918;
Best Local Similarity 93.3%; Pred. No. 7.1e-139;
Matches 571; Conservative 0; Mismatches 33; Indels 8; Gaps 5;

Qy      1 ATGTGTCTCAGTACTGCTCTCTTGTGATGTTGAAAACCTCATATCAACATTCAA 60
      |||||
Db      308 ATGTGTCTCAGTACTGCTCTCTTGTGATGTTGAAAACCTCATATCAACATTCAA 367
      |||||

Qy      61 GACTCTATCACCTCACCCAGTGTCTTAACAAGCTCCATTTTGAATCAATAGCAGGCGC 120
      |||||
Db      368 GACTCTATCACCTCACCCAGTGTCTTAACAAGCTCCATTTTGAATCAATAGCAGGCGC 427
      |||||

Qy      121 ATATCTTTTCTCAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAGATTTCAGGTGCC 180
      |||||
Db      428 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAGATTTCAGGTGCC 487
      |||||

Qy      181 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGAGCCCAAGCCCTAGTT 240
      |||||
Db      488 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGAGCCCAAGCCCTAGTT 547
      |||||

Qy      241 ACTCACAGCAGGGAACCACTGCCAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATT 300
      |||||
Db      548 ACTCACAGCAGGGAACCACTGCCAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATT 607
      |||||

301 CTGTGTTACATTTGTTGTTCCCAACAGCTCCCAACTGCAAGAACTGCAATTCAGGCC 360
      |||||
608 CTGTGTTACATTTGTTGTTCCCAACAGCTCCCAACTGCAAGAACTGCAATTCAGGCC 667
      |||||

361 TATGGAGCATCGATAGTACTGTGACCAAGTGTGAGGAGTCCAGAGAAAGGTCACTCAA 420
      |||||
668 TATGGAGCATCGATAGTACTGTGACCAAGTGTGAGGAGTCCAGAGAAAGGTCACTCAA 726
      |||||

421 AGAATTTATCAAGAAACAGAGGCTATCTTGGTCCATCCCAACAGGAGCTGCAATTCAA 480
      |||||
727 AGAATTTATCAAGAAACAGAGGCTATCTTGGTCCATCCCAACAGGAGCTGCAATTCAA 786
      |||||

481 GCTGACAAAGGAACAATTCCTGCTGAAAGTGTCTGAACAGGTTCCCTTGG-TAGATGCACT 539
      |||||
787 GCTGACAAAGGAACAATTCCTGCTGAAAGTGTCTGAACAGGTTCCCTTGGNTAGATGCACT 846
      |||||

540 -GCTGGTACCAGTACG--AGGAGGAGGAGTGTGCTGGAATA--GCCATTACATTTAA 593
      |||||
847 GGTGGTACCAAGTAGGAGGAGGAGTGTGCTGGAATAAGCCATTTACATTTAA 906
      |||||

594 GGCCCTGAAACC 605
      |||||
907 GGCCCTGGAACC 918
      |||||

RESULT 10
BI738966      823 bp      mRNA      linear      EST 20-SEP-2001
LOCUS      603359780F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5367043 5',
DEFINITION      mRNA sequence.
ACCESSION      BI738966
VERSION      BI738966.1 GI:15715979
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
      house mouse.

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-remail.nih.gov
      Tissue Procurement: The Cepko Laboratory
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM1933 row: p column: 20
      High quality sequence stop: 799.

FEATURES
      Location/Qualifiers
          1..823
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="IMAGE:5367043"
              /clone_lib="NIH_MGC_94"
              /tissue_type="retina"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 3.3 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH_MGC Library."
          237 a      208 c      193 g      184 t      1 others
BASE COUNT      237 a      208 c      193 g      184 t      1 others
ORIGIN
Query Match      47.1%; Score 479.8; DB 13; Length 823;
Best Local Similarity 98.9%; Pred. No. 6.2e-134;
Matches 525; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy      1 ATGTGTCTCAGTACTGCTCTCTTGTGATGTTGAAAACCTCATATCAACATTCAA 60
      |||||

```

```

Db      274 ATGTGCTCAGTACTCATCTCTCTGCTGATGTTGAAAAGCTCATATCAACATTCAA 333
Qy      61  GACTCTATCCACTCACCAGGCTGCTAACAAGCTTCCATTTTGAATCAAAATAGCAGGCGC 120
Db      334 GACTCTATCCACTCACCAGGCTGCTAACAAGCTTCCATTTTGAATCAAAATAGCAGGCGC 393
Qy      121 AATCTTTTCTTCAAAATGAGCTTCTCCAGAAACTGGGCTTTTAAAGATTCGAGGTGCC 180
Db      394 AATCTTTTCTTCAAAATGAGCTTCTCCAGAAACTGGGCTTTTAAAGATTCGAGGTGCC 453
Qy      181 CTTAATGCCATCAGAGCTTAATCTCTGACAGCCGAGAGAGAGCCCAAGCCGTAGTT 240
Db      454 CTTAATGCCATCAGAGCTTAATCTCTGACAGCCGAGAGAGAGCCCAAGCCGTAGTT 513
Qy      241 ACTCACAGCAGCGGAAACCATG-GCCAAAGCTCTCACCTATGCTGCTAACTGGAAGGAAT 299
Db      514 ACTCACAGCAGCGGAAACCATGNGCCAAAGCTCTCACCTATGCTGCTAACTGGAAGGAAT 573
Qy      300 TCCTGCTTACATGTTGGTTCCTCCAAACAGCTCCCAACTGCAAG-AAACTGGCAATCCAAG 358
Db      574 TCCTGCTTACATGTTGGTTCCTCCAAACAGCTCCCAACTGCAAGAAACTGGCAATCCAAG 633
Qy      359 CCTATGAGCATCGATAGTATAC-TGTGACCCCAAGTCAAGTCCAGAGAAAGGTCACT 417
Db      634 CCTATGAGCATCGATAGTATAC-TGTGACCCCAAGTCAAGTCCAGAGAAAGGTCACT 693
Qy      418 CAAAGAAATTTGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAAGT 477
Db      594 CAAAGAAATTTGCG-AGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAAGT 752
Qy      478 ATAGCTGGACAGGAACAATGGCCCTGGAAAGTCTGAACAGGTTCCTCTTG 528
Db      753 ATAGCTGGACAGGAACAATGGCCCTGGAAAGTCTGAACAGGTTCCTCTTG 803

RESULT 11
AV022510 698 bp mRNA linear EST 23-OCT-2001
LOCUS AV022510 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
DEFINITION clone 1190028F16, mRNA sequence.
ACCESSION AV022510
VERSION AV022510.2 GI:16356391
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 698)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyama,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On May 11, 1999 this sequence version replaced gi:4799502.
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@tc.riken.go.jp
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

```

```

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
Location/Qualifiers
1..698
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1190028F16"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"
/sex="mixed"
/dev_stage="18-day embryo"
/note="Organ: mammary gland; Vector: pT7T3D-pac (Pharmacia
RI: 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAATGGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTTTTTTTTTT
T 3'] ; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library.
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 203 a 150 c 171 g 174 t
ORIGIN
Query Match 46.7%; Score 475.2; DB 9; Length 698;
Best Local Similarity 98.4%; Pred. No. 1.4e-132;
Matches 480; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 529 GTAGATGCACCTGGTGTACCAGTAGGAGGAGGAGGAGTGGTGTGCGTAATAGCCATTACA 588
Db 1 GTAGATGCACCTGGTGTACCAGTAGGAGGAGGAGGAGTGGTGTGCGTAATAGCCATTACA 60
Qy 589 ATTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTAGCCCTCGAATGCGAGATGAC 648
Db 61 TTTAAGGCCCTGAAACCTAGTGTGAAGGTATACGCTGTAGCCCTCGAATGCGAGATGAC 120
Qy 649 TGCTACCACTGCTAAACTGAAAGGAGAACTGACCCCAATCTTCATCTCCCAAGAACCATTA 708
Db 121 TGCTACCACTGCTAAACTGAAAGGAGAACTGACCCCAATCTTCATCTCCCAAGAACCATTA 180
Qy 709 GCAGATGGTGTCAAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTG 768
Db 181 GCAGATGGTGTCAAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTG 240
Qy 769 GATGATGCTTCTACTGTCCAGGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGG 828
Db 241 GATGATGCTTCTACTGTCCAGGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGG 300
Qy 829 AGAATGAACCTGCTCATTGAGCCGACCTGCTGGCCGTGGCAGCTGGCTGCTCTCAG 888
Db 301 AGAATGAACCTGCTCATTGAGCCGACCTGCTGGCCGTGGCAGCTGGCTGCTCTCAG 360

```

Qy 889 CATTTCACAAACAGTCTCTCCAGAAAGTAAGAACCTGTCATTTGACTCAGTGGGGGAAT 948  
 Db 361 CATTTCACAAACAGTCTCTCCAGAAAGTAAGAACCTGTCATTTGACTCAGTGGGGGAAT 420  
 Qy 949 GTACACCTAACCTCCCTGAACTGGTGGGCGAGCGCTGAAGCGGCAGCTCTTACCAGAGC 1008  
 Db 421 GTACACCTAACCTCCCTGAACTGGTGGGCGAGCGCTGAAGCGGCAGCTCTTACCAGAGC 480  
 Qy 1009 GTCTGTTT 1016  
 Db 481 GTTCTGT 488

## RESULT 12

AV374491

LOCUS

AV374491 602 bp mRNA linear EST 24-OCT-2001  
 cDNA clone 9130011P22 3', mRNA sequence.

ACCESSION

AV374491

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 602)  
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
 Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,  
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,  
 D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
 Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

COMMENT

On Nov 14, 1999 this sequence version replaced gi:6422138.  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

M., Konno, H., Okazaki, Y., Muramatsu, M., Sugahara, Y., Shibata, K., Itoh  
 Carninci, P., Shibata, Y., Hayashi, N., Sugahara, Y., Shibata, K., Itoh  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
 Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 X. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to

## FEATURES

source

prepare mouse tissues.

Location/Qualifiers

1..602

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="9130011P22"

/clone\_lib="RIKEN full-length enriched, adult male cecum"

/sex="male"

/tissue\_type="cecum"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization

to Rot = 10.0 and subtraction to Rot = 185.0. Second

strand cDNA was prepared with the primer adapter of

sequence [5' GAGAGAGAGATCTCGAGTTAATAATTAATCCCCCCCCCC

3']. cDNA was cloned into the XhoI and BamHI sites.

vector: a modified pBluescript KS(+) after bulk excision

from lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

BamHI"

BASE COUNT 167 a 133 c 148 g 152 t 2 others.

ORIGIN

Query Match 45.3%; Score 461.2; DB 10; Length 602;  
 Best Local Similarity 94.7%; Pred. No. 2.2e-128;  
 Matches 487; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

Qy 503 TGGAAAGTCTGAACACAGTTCCCTTGTAGATGCACCTGGTGGTACCATAGGAGGAGAG 562

Db 1 TTGAAGTCTGAACACAGTTCCCTTGTAGATGCACCTGGTGGTACCATAGGAGGAGAG 58

Qy 563 GAATGGTTGCTGGAATAGCCATTACAATTAAGGCCCTTGAACCTAGTGTGAAGGTATACG 622

Db 59 GAATGGTTGCTGGAATAGCCATTACAATTAAGGCCCTTGAACCTAGTGTGAAGGTATACG 118

Qy 623 CTGCTGAGCCCTCGAATGCAGATGCTGTACCACTTAACTGAAAGGAGAACTGACCC 682

Db 119 CTGCTGAGCCCTCGAATGCAGATGCTGTACCACTTAACTGAAAGGAGAACTGACCC 178

Qy 683 CCAATCTTCATCTCCAGAAACCATACAGATGGTGTCAATCCAGCATTTGGCTTGAATA 742

Db 179 CCAATCTTCATCTCCAGAAACCATACAGATGGTGTCAATCCAGCATTTGGCTTGAATA 238

Qy 743 CCTGGCCCTATTATAAGAGACCTTTGTGGATGATGTCTTCACTGTCCACGAAAGTAAATCA 802

Db 239 CCTGGCCCTATTATAAGAGACCTTTGTGGATGATGTCTTCACTGTCCACGAAAGTAAATCA 298

Qy 803 AGTATGCAACCCAGCTGGTGTGGGGGAGAAATGAACCTGCTCATTTGACCCGAGCTGCTGGC 862

Db 299 AGTATGCAACCCAGCTGGTGTGGGGGAGAAATGAACCTGCTCATTTGACCCGAGCTGCTGGC 358

Qy 863 TGGCACTGGCTGCAGTCTCTCTCAGCATTTTCAACACAGTCTCTCCAGAAAGTAAAGAACG 922

Db 359 TGGCACTGGCTGCAGTCTCTCTCAGCATTTTCAACACAGTCTCTCCAGAAAGTAAAGAACG 418

Qy 923 TCTGCATTTGACTCAGTGGGGGAAATGATAGACCTTAACCTCCCTGAACTGGTGGGGCAGG 982

Db 419 TCTGCATTTGACTCAGTGGGGGAAATGATAGACCTTAACCTCCCTGAACTGGTGGGGCAGG 478

Qy 983 CTGAACGGCCAGCTCTCTTACCACAGCGTCTGTTT 1016

Db 479 CTGAACGGCCAGCTCTCTTACCACAGCGTCTGTTT 512



```
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 454.
Location/Qualifiers
I. .455
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="IMAGE:467086"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCGAATGGAGCGCGCGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT 140 a 121 c 90 g 104 t
ORIGIN
Query Match 44.7%; Score 455; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.4e-126;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 CATCTCCTTTGTGATGTTGAAAAAGCTCATATCAACATTCAGACTCTATGCACCTCAC 77
Db 1 CATCTCCTTTGTGATGTTGAAAAAGCTCATATCAACATTCAGACTCTATGCACCTCAC 60
QY 78 CCAGTGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTCAAAATG 137
Db 61 CCAGTGTCTAACAAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTCAAAATG 120
QY 138 TGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTCGAGGTGGCCCTTAATGCCATCAGAG 197
Db 121 TGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTCGAGGTGGCCCTTAATGCCATCAGAG 180
QY 198 CTTAATTCCTGCACGCCAGAGAGAGCCAAAGCCGTAGTTACTCAGCAGCGCGAAA 257
Db 181 CTTAATTCCTGCACGCCAGAGAGAGCCAAAGCCGTAGTTACTCAGCAGCGCGAAA 240
QY 258 CCATGGCCCAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGGT 317
Db 241 CCATGGCCCAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGGT 300
QY 318 TCCCAAAACAGCTCCCAACTGCAAGAACTGCGAATCCAAAGCCTATGGAGCATCGATAGT 377
Db 301 TCCCAAAACAGCTCCCAACTGCAAGAACTGCGAATCCAAAGCCTATGGAGCATCGATAGT 360
QY 378 ATACTGTGACCAAGTGCAGCTCCAGAGAAAGCTCACTCAAGAAATTTATCAAGAAAC 437
Db 361 ATACTGTGACCAAGTGCAGCTCCAGAGAAAGCTCACTCAAGAAATTTATCAAGAAAC 420
QY 438 AGAAGGCATCTTGGTCCATCCCAACCCAGAGGCGTG 472
Db 421 AGAAGGCATCTTGGTCCATCCCAACCCAGAGGCGTG 455
RESULT 15
BG965678 729 bp mRNA linear EST 12-JUN-2001
LOCUS 602830522F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4985237 5',
DEFINITION mRNA sequence.
ACCESSION BG965678
VERSION BG965678.1 GI:14353315
```

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 729)  
AUTHORS NIH-MGC http://imgc.ncbi.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue procurement: Jeffrey E. Green, M.D.  
CDNA Library preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10992 row: 1 column: 06  
High quality sequence stop: 719.  
Location/Qualifiers  
I. .729  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:4985237"  
/clone\_lib="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: colon; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 211 a 163 c 174 g 181 t  
ORIGIN  
Query Match 43.8%; Score 445.6; DB 13; Length 729;  
Best Local Similarity 98.9%; Pred. No. 1.4e-123;  
Matches 459; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 553 GGAGGAGGAGGAATGGTTCCTGGAATAGCCATACAAATTAAGCCCTGAACCTAGTGTG 612  
Db 1 GGAGGAGGAGGAATGGTTCCTGGAATAGCCATACAAATTAAGCCCTGAACCTAGTGTG 60  
QY 613 AAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGTACAGTCTAACTGAAGGA 672  
Db 61 AAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGTACAGTCTAACTGAAGGA 120  
QY 673 GAAC TGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCAT 732  
Db 121 GAAC TGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCAT 180  
QY 733 GGCTTGAATACCTGGCCCTATTATAAGAGACCTTGGATGATGTCTTCACTGTCCAGGA 792  
Db 181 GGCTTGAATACCTGGCCCTATTATAAGAGACCTTGGATGATGTCTTCACTGTCCAGGA 240  
QY 793 GATGAATCAAGTATGCAACCCAGCTGGTGGGGGAGAAATGAACTGCTCAATGAGCG 852  
Db 241 GATGAATCAAGTATGCAACCCAGCTGGTGGGGGAGAAATGAACTGCTCAATGAGCG 300  
QY 853 ACTGCTGGCGTGGCACTGGCTGCAGTGTCTTCAGCAATTTCCAAACAGTCTCTCCAGAA 912  
Db 301 ACTGCTGGCGTGGCACTGGCTGCAGTGTCTTCAGCAATTTCCAAACAGTCTCTCCAGAA 360  
QY 913 GTAAGAACGCTGTGCTTACTAGTGGGGGGAATGTAGACCTTACCTCCCTGAACTGG 972  
Db 361 GTAAGAACGCTGTGCTTACTAGTGGGGGGAATGTAGACCTTACCTCCCTGAACTGG 420  
QY 973 GTGGGCGAGCTGAACGGCCAGCTCTTACCAGAGCGTCTGTT 1016  
Db 421 GTGGGCGAGCTGAACGG-CAGCTCTTACCAGAGCGTCTGTT 463

Search completed: June 24, 2003, 05:35:27



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:36:57 ; Search time 19.4713 Seconds  
(without alignments)  
722.114 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISFADVEKAHINIQ.....TSLNMGQAERPAPYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1740	100.0	339	1 SRR_MOUSE	O9qzx7 mus musculus
2	1582.5	90.9	340	1 SRR_HUMAN	O9qzta homo sapien
3	587	33.7	326	1 YK8_YEAST	P36007 saccharomyc
4	560	32.2	323	1 YCNE_SCHPO	O59791 schizosacch
5	491	28.2	329	1 THD2_SALTY	PI1954 salmonella
6	476	27.4	329	1 THD2_ECOLI	P05792 escherichia
7	436	25.1	332	1 Y4TJ_RHISN	P55664 rhizobium s
8	421	24.2	513	1 THD1_PASMU	Q9ckj2 pasteurella
9	405	23.3	513	1 THD1_HAEIN	P46493 haemophilus
10	395.5	22.7	576	1 THDH_YEAST	P00927 saccharomyc
11	384	22.1	514	1 THD1_SALTY	P20506 salmonella
12	379.5	21.8	550	1 THDH_ARKAD	O42615 arxula aden
13	378	21.7	514	1 THD1_ECOLI	O4968 escherichia
14	371.5	21.4	595	1 THD1_LYCES	P25306 lycopersico
15	371	21.3	592	1 THD1_ARKAD	O92856 arabidopsis
16	364	20.9	422	1 THD1_BACSU	P37946 bacillus su
17	362.5	20.8	416	1 THD1_LACLA	O02145 lactococcus
18	358	20.6	590	1 THD1_CICAR	O39469 cicier ariet
19	352	20.2	507	1 THD1_BURCE	P53607 burkholderi
20	339	19.5	415	1 THD1_BACHD	Q9kc83 bacillus ha
21	323.5	18.6	429	1 THD1_MYCTU	O10766 mycobacteri
22	307	17.6	436	1 THD1_CORGL	O04513 corynebacte
23	300.5	17.3	427	1 THD1_MYCLE	O9x7f1 mycobacteri
24	283.5	16.3	362	1 SDHL_RAT	O9x7f1 rattus norv
25	252.5	14.5	328	1 SDHL_HUMAN	P20132 homo sapien
26	218.5	12.6	359	1 THD1_SOLTU	P31212 solanum tub
27	212	12.2	352	1 THRC_BACSP	P09123 bacillus sp
28	208.5	12.0	352	1 THRC_BACSU	P04990 bacillus su
29	208	12.0	312	1 CYSK_SYNY3	P73410 synecocyst
30	202.5	11.6	311	1 CYSK_BACSU	O34476 bacillus su
31	201	11.6	307	1 CYSK_BACSU	P37887 bacillus su
32	200	11.5	338	1 SDHL_YEAST	P17324 saccharomyc
33	198.5	11.4	310	1 CYSK_MYCLE	O32978 mycobacteri

34	198.5	11.4	310	1	CYSK_MYCTU	P95230 mycobacteri
35	198.5	11.4	382	1	THRC_SYNY3	P74193 synecocyst
36	189.5	10.9	307	1	CYSK_FLASP	O59447 flavobacter
37	187.5	10.8	354	1	THRC_BACHD	O9k7e3 bacillus ha
38	185.5	10.7	550	1	CBS_HUMAN	P35520 homo sapien
39	181	10.4	322	1	CYSK_ARKAD	P47998 arabidopsis
40	177	10.2	322	1	CYSK_BRAJU	O32733 brassica ju
41	177	10.2	325	1	CYSK_CITLA	O43317 citrullus l
42	176	10.1	299	1	CYSM_CAMJE	P71128 campylobact
43	176	10.1	560	1	CBS_RAT	P32232 rattus norv
44	172	9.9	324	1	CYSK2_BRAJU	O32735 brassica ju
45	172	9.9	325	1	CYSK2_ORVSA	O9xeab oryza sativ

ALIGNMENTS

RESULT 1  
SRR\_MOUSE STANDARD; PRT; 339 AA.  
AC O9QZX7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine racemase (EC 5.1.1.-).  
GN SRR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RX MEDLINE=20027561; PubMed=10557334;  
RA Wolosker H., Blackshaw S., Snyder S.H.;  
RT "Serine racemase: a glial enzyme synthesizing D-serine to regulate  
glutamate-N-methyl-D-aspartate neurotransmission.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:13409-13414(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.  
CC -!- TISSUE SPECIFICITY: Brain.  
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF148321; AAF08701.1;  
DR EMBL; BC011164; AAH11164.1;  
DR HSSP; P04968; ITDJ  
DR MGD; MGI:1351636; Srr.  
DR InterPro; IPR001926; B6\_enzyme\_beta.  
DR InterPro; IPR000634; S/T\_dehydratse.  
DR Pfam; PF00291; PALP; 1.  
DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
KW Isomerase; Pyridoxal phosphate.  
FT BINDING 56  
SQ SEQUENCE 339 AA; 36359 MW; B9AE9A933658728 CRC64;  
PYRIDOXAL PHOSPHATE (BY SIMILARITY).

Query Match 100.0%; Score 1740; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 6.5e-126;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCAQYCISFADVEKAHINIQSDIHLTPVLTSILNQIAGRNLFKCELFQKTSFKIRGA 60  
|||||

Db 1 MCAQYCISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLPFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIA 120  
Db 61 LNAIRGLIPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIA 120  
QY 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNOEPVAVIAGQGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNOEPVAVIAGQGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLHPPTIADGV 240  
Db 181 VVPVGGGMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLHPPTIADGV 240  
QY 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVAAVLSQHFQT 300  
Db 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVAAVLSQHFQT 300  
QY 301 VSPVKNVCIVLSGNNVLDLTLNMGVGAERPAPYQTVSV 339  
Db 301 VSPVKNVCIVLSGNNVLDLTLNMGVGAERPAPYQTVSV 339

## RESULT 2

SRR\_HUMAN  
ID SRR\_HUMAN STANDARD; PRT; 340 AA.  
AC Q9GZT4;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine racemase (EC 5.1.1.-).  
GN SRR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20510003; PubMed=11054547;  
RA Demiranda J., Santoro A., Engelender S., Wolosker H.;  
RT "Human serine racemase: molecular cloning, genomic organization and  
functional expression";  
RL Gene 256:183-188(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murekawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDD human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.  
CC -!- TISSUE SPECIFICITY: Brain. Expressed at high levels in hippocampus  
CC and corpus callosum, intermediate levels in substantia nigra and  
CC caudate, and low levels in amygdala, thalamus, and subthalamic  
CC nuclei.  
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF169974; AAC27081.1;  
CC DR EMBL: AK023169; BAB14442.1;  
CC DR Genbank: HGNC:14398; SRR.

DR MIM; 606477;  
DR InterPro; IPR001926; B6\_enzyme\_beta.  
DR InterPro; IPR00634; S/T\_dehydrtse.  
DR Pfam; PF00291; PALP; 1.  
DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
KW isomerase; Pyridoxal phosphate.  
FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 340 AA; 36566 MW; 873342C62D5D7B9D CRC64;  
  
Query Match 90.98; Score 1582.5; DB 1; Length 340;  
Best Local Similarity 89.78; Pred. No. 7.2e-114;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
  
QY 1 MCAQYCISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLPFKCELFQKTSFKIRGA 60  
Db 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQIAGRNLPFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIA 120  
Db 61 LNAIRGLIPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVPTAPNCKKLAIA 120  
QY 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNOEPVAVIAGQGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCDPSDESREKVTQIMQETGILVHPNOEPVAVIAGQGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLHPPTIADGV 240  
Db 181 VVPVGGGMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLGELTPNLHPPTIADGV 240  
QY 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVAAVLSQHFQT 300  
Db 241 KSSIGLNTWPIIRDLDVDFVTVEDEIKYATQLVWGRMKLLIEPTAGVAAVLSQHFQT 300  
QY 301 VSPVKNVCIVLSGNNVLDLTLNMGVGAERPAPYQTVSV 339  
Db 301 VSPVKNVCIVLSGNNVLDLTLNMGVGAERPAPYQTVSV 339

## RESULT 3

YKVB\_YEAST  
ID YKVB\_YEAST STANDARD; PRT; 326 AA.  
AC P36007;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical 34.9 kDa protein in COS9-JEN1 intergenic region.  
GN YKL218C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=95028164; PubMed=7941750;  
RA Tzeremia M., Horaitis O., Alexandraki D.;  
RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI  
RT identified the known loci URA1, SAC1 and TRP3, and revealed 6 new  
RT open reading frames including homologues to the threonine  
RT dehydratases, membrane transporters, hydantoinsases and the  
RT phospholipase A2-activating protein.";  
RL Yeast 10:663-679(1994).  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC

```

DR EMBL; X75951; CAA53555.1; -
DR EMBL; Z28218; CAA82063.1; -
DR PIR; S38061; S38061.
DR PIR; S44320; S44320.
DR HSP; P04968; ITDJ.
DR SGD; S0001701; YKL218C.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).
FT BINDING 53 53
SQ SEQUENCE 326 AA; 34899 MW; 21CF7EFC8AB4431 CRC64;

Query Match 33.7%; Score 587; DB 1; Length 326;
Best Local Similarity 40.9%; Pred. No. 8.e-38;
Matches 128; Conservative 66; Mismatches 111; Indels 8; Gaps 4;

QY 8 SFADVEKAHINIQDSIHLTPVLTSILNQTAGRNLFKCELFQKTSFKIRGALNAIRGL 67
   : : || | : : : ||||| : || | : : || | : : || | : : || | : : || | :
Db 5 TYGDVLDSNRKIEYVKNTPVLTSLMLNDRLGAQIVFKGENFQVGAFFRGAMNAVSKL 64

QY 68 IPTPEKPKAVYTHSSGNHGQALTYAAKLEGIPAVIVPOTAPNCKKLAIOAYGASIVY 127
   : : : : : : ||||| : : : : : : ||||| : : : : : : ||||| : : : : : :
Db 65 --SDEKRSKGVITAFSSGNHQAIALSAKLLNVPATIVMPEDAPALKVAATAGYGAHIR 121

QY 128 CDPDSREKVTQRIQMOETEGILVHPNQEPAVIAGOGTIALEVLNOVPLVDALVWPVGG 187
   : : : : : : ||||| : : : : : : ||||| : : : : : : ||||| : : : : : :
Db 122 YNYTEDREQIGRLAAEHGFALIPYDHPDVIAGOGTSKELLEVGQDLALFVPLGG 181

QY 188 GWVAGIATITKALKPSVYVAAPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246
   : : : : : : ||||| : : : : : : ||||| : : : : : : ||||| : : : : : :
Db 182 GLLSGSALAARSLSPCKIFGVEPEAGNDGQSFSGSIV-HINTEKTIADGAQTQHLGE 240

QY 247 NTPPIIRDLVDVFTTDEIKATQLVGRMKLLIEPTAGVALAAVLSOHFTQVSPEVK 306
   : : : : : : ||||| : : : : : : ||||| : : : : : : ||||| : : : : : :
Db 241 YPAIRIIRDLVDVFTTDEIKATQLVGRMKLLIEPTAGVALAAVLSOHFTQVSPEVK 306

QY 307 NVCIVLSSGGNVDL 319
   : : : : : : ||||| : : : : : : ||||| : : : : : : ||||| : : : : : :
Db 298 KVGIISSGGNVDM 310

RESULT 4
YCNE_SCHPO STANDARD; PRT; 323 AA.
AC O59791;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein C320.14 in chromosome III.
GN SPCC320.14 OR SPCC330.15C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sources J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

```

GN TDCB OR STM3244.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RN SEQUENCE OF 1-24.  
 RX MEDLINE=83023208; PubMed=6751404;  
 RA Kim S.S., Datta P.;  
 RT "Chemical characterization of biodegradative threonine dehydratases  
 from two enteric bacteria";  
 RL Biochim. Biophys. Acta 706:27-35(1982).  
 CC -|- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 H(2)O.  
 CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -|- ENZYME REGULATION: EACH PROTEIN MOLECULE CAN BIND UP TO FOUR  
 CC MOLECULES OF AMP, WHICH ACTS AS AN ALLOSTERIC ACTIVATOR TO THE  
 CC ENZYME. THE ENZYME IS ALSO INHIBITED BY ALPHA-KETO ACIDS AND OTHER  
 CC CATABOLITES. IT IS ALLOSTERICALLY INHIBITED BY ISOLEUCINE  
 CC AND ALLOSTERICALLY ACTIVATED BY VALINE.  
 CC -|- PATHWAY: Threonine catabolism.  
 CC -|- SUBUNIT: HOMOTETRAMER.  
 CC -|- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE008849; AAL22117.1;  
 DR PIR; B22317; B22317.  
 DR StyGene; SG10390; tdcB.  
 DR InterPro; IPR001926; B6\_enzyme\_beta.  
 DR InterPro; IPR00634; S/T\_dehydrtse.  
 DR Pfam; PF00291; PALP; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 DR KW Lyase; Pyridoxal phosphate; Allosteric enzyme; Complete proteome.  
 FT BINDING 58 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 24 24 K -> I (IN REF. 2).  
 FT SEQUENCE 329 AA; 35141 MW; 1C619B021DE817C CRC64;  
 Query Match  
 Best Local Similarity 36.8%; Score 491; DB 1; Length 329;  
 Matches 121; Conservative 68; Mismatches 116; Indels 24; Gaps 8;  
 QY 7 ISADVEKAHINTQSDTHLTPVLTSSLTNIOAGRNLFKCELTQKTSFKIRGALNAIRG 66  
 Db 9 VAIEDILEAKKRLAGIKYTKTGMPSRNSYFSERCKGEIFLKFENMORTGSGFKIRGAFNKLSS 68  
 QY 67 LIPDTPEKPAVVTHSSNGHGAALTYAAKLEGIPAVIVVPOTPAPNCKLAIQYGASIV 126  
 Db 69 L---TEAKKRGVVACSGAGHAGVLSLSCAMLGIDGKVVMPKAPKSKVAATCDYSAEVV 125  
 QY 127 -YCDPSSDEKRYQIMQETEG-ILVHPNQEPVIAVGQGTIALEVLNQVPLVDALVVPV 184  
 Db 126 LHGDNFNDFITAKYSEIV--ETEGRIPTFPYDDPKPIAGQGTIGLEIMEDLYDNDVIVPI 183  
 QY 185 GGGGWWAGIAITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK-SS 243

## RESULT 6

THD2\_ECOLI  
 ID THD2\_ECOLI STANDARD; PRT; 329 AA.  
 AC P05792;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).  
 GN TDCB OR B3117 OR Z4469 OR ECS3997.  
 OS Escherichia coli, and  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=87092415; PubMed=3540965;  
 RA Datta P., Goss T.J., Omnaas J.R., Patil R.V.;  
 RT "Covalent structure of biodegradative threonine dehydratase of  
 RT Escherichia coli: homology with other dehydratases";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=89282418; PubMed=2660107;  
 RA Schweizer H., Datta P.;  
 RT "The complete nucleotide sequence of the tdc region of Escherichia  
 RT coli.";  
 RL Nucleic Acids Res. 17:3994-3994(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Tada T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli



```

QY 241 KSSTGL-NTW--PIIRDLVDVFTTEDEIKYATQLVWGRMKLLIEPTAGVALAALVSOH 297
DB 237 GGGIGMANWTFQMCRRALLDVLNVEGEITAAGIRHAYEHQRILEGAGAVGIAULLS-- 294
QY 298 FQTVSPVKNVCIVLSGNDVL 319
DB 295 -GKYAARGSGVGVLSQNDIM 315

RESULT 8
THDI_PASMU STANDARD; PRT; 513 AA.
AC Q9CKJ2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
DE deaminase).
GN ILVA OR PM1624.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
OC Pasteurella
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
CC threonine in a two-step reaction. The first step is a dehydration
CC of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
CC H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006199; AAK03708.1; -
CC HSSP; P04968; ITDJ.
CC InterPro; IPR001926; B6_enzyme_beta.
CC InterPro; IPR000634; S/T_dehydrtse.
CC InterPro; IPR001721; ThrDb_C.
CC Pfam; PF00291; PALP; 1.
CC Pfam; PF00585; Thr_dehydrat_C; 2.
CC TIGRfams; TIGR01124; ilva_2Cterm; 1.
CC PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
CC Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
CC Complete proteome.
CC BINDING 61
CC SEQUENCE 513 AA; 56288 MW; 3E427ADC54E5FC7C CRC64;

Query Match 24.2%; Score 421; DB 1; Length 513;
Best Local Similarity 34.1%; Pred. No. 7.7e-25;
Matches 106; Conservative 63; Mismatches 134; Indels 8; Gaps 6;

QY 19 IQDSIHLPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNIRGLIPTPEPKPA 78
DB 24 VYDVAQVTPQDMAKLSRLGNKVFIKREDQPVHFSFKLRGAYAMIAGL---SAEQKASG 80
QY 79 VYTHSSGNHQAITYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGASIVYCDPS-DESRK 137
DB 81 VTAASAGNAQGVALSAAKHLGRALIVMPONTPSIKVDVAVRGFGVLLHGAFNDEAKAK 140

```

```

QY 138 VTORIMQETGILVHPNQPAPVITAGOGTIALEVLNOVPLVDALVVPVGGGGMVAGIAITI 197
DB 141 AIE-LAESKNWTFIPFDPFAVITAGOGSTAMELLOQNSQIDRIFVPVGGGLAAGIAVLI 199
QY 198 KALKPSVKVYAARPSNADDCYQSKLKGELTPNLPHPETIADGVK-SSIGLNTWPIIRDLV 256
DB 200 KQLMPEIKVIGVE-SKDSACLRYALKRAGKPIDLDRVGLFADGVAVKRGIDETFRVCCQYI 258
QY 257 DVVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAALVSOHFOFOTVSPVKNVCIVLSGNN 316
DB 259 DDVVLVDGDEICAAVKDIFENWRAIAEPSGALSAG-LKKYVREHNTQGETTLNVLSGAN 317
QY 317 VDLTSLNWWGQ 327
DB 318 LNFHTLRYVSE 328

RESULT 9
THDI_HAEIN STANDARD; PRT; 513 AA.
AC P46493;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
DE deaminase).
GN ILVA OR HI0738.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen B.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RA Koonin E.V., Rudd K.E.;
RL Submitted (SEP-1995) to the SWISS-PROT data bank.
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
CC threonine in a two-step reaction. The first step is a dehydration
CC of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
CC H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32757; AAC22398.1; -
CC HSSP; P04968; ITDJ.
CC TIGR; HI0738.1; -

```

```

DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydratase.
DR InterPro: IPR001721; ThrDh_C.
DR Pfam: PF00291; PALP.
DR Pfam: PF00585; Thr_dehydrat_C_2.
DR TIGRfams: TIGR01124; liva_2Cterm; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
FT Complete proteome.
FT BINDING 63
SQ SEQUENCE 513 AA; 56662 MW; DF42CA8B6FDE4CD7 CRC64;

Query Match 23.3%; Score 405; DB 1; Length 513;
Best Local Similarity 32.6%; Pred. No. 1.3e-237;
Matches 104; Conservative 68; Mismatches 13; Indels 10; Gaps 8;

QY 12 VEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNATRGILPDT 71
DB 19 IVKLSRVYEAQVTPLOKMGKLSERLHNNIWKREDRQPVNFEKLGAYAMISL---$ 75
QY 72 PEEKRAVTHSSGNHGQALTYAAKLEGIPAYIVPQTAPNCKKLAIQAYGASIVYCDPS 131
DB 76 AEQKAAGVIAAGNHAQGVAAKQGLKALIVMPQNTPSIKVDVAVRGFGEVLLHGAN 135
QY 132 -DESREKVTQRIQETEGILVHNPQPAVITAGOGTTALEVLNQLVDALVVPVGGGMV 190
DB 136 FDEAKAKAIE-LSKEKMTTIPFDHPLVITAGOGTTALEMLOQVADLDYVFGVGGGLA 194
QY 191 AGIATITKALKPSVKVYAAEPSNADDCYQSKL-KGELTPNLHPPTIADGVK-SSICLNT 248
DB 195 AGVAILLKQMPKIKIGV-BKDSACLKALOKGEPDTH-IGLPADGVAVKRGDET 252
QY 249 WPIRLDLDVFTVTEDEIKYATQLVGRMKLLIETAGVALAALVLSQHFQTVSPVKVY 308
DB 253 FRLCQQLDMDLVDSDEVAAMKDLFENVRAEPPSGALGLAG-LKKYVKQNHIEGKNM 311
QY 309 CIVLSGGNVDLTSLNNVGO 327
DB 312 AAILSGANLNFHTLRVYSE 330

RESULT 10
THDH_YEAST STANDARD; PRT; 576 AA.
AC P00927;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase, mitochondrial precursor (EC 4.2.1.16)
DE (Threonine deaminase).
DE ILV1 OR YER086W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Kiehlbrand-Nandt M.C., Holmberg S., Petersen J.G.L.,
RA Nilsson-Tillgren T.;
RA "Nucleotide sequence of the gene for threonine deaminase (ILV1) of
RT Saccharomyces cerevisiae."
RL Carlsberg Res. Commun. 49:567-575(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=528C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
CC H(2)O.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC -1- PATHWAY: Isoleucine biosynthesis; first step.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M36383; AAA34705.1; -.
CC EMBL: X01466; CAA25696.1; -.
CC EMBL: U18839; AAB64641.1; -.
CC PIR: A01150; DMBYT.
CC HSP: P04968; ITDJ.
CC SGP: S0000888; ILV1.
CC InterPro: IPR001926; B6_enzyme_beta.
CC InterPro: IPR000634; S/T_dehydratase.
CC InterPro: IPR001721; ThrDh_C.
CC Pfam: PF00291; PALP.
CC Pfam: PF00585; Thr_dehydrat_C_2.
CC TIGRfams: TIGR01124; liva_2Cterm; 1.
CC PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
KW Isoleucine biosynthesis; Lyase; Mitochondrion; Pyridoxal phosphate;
KW Transit peptide; Allosteric enzyme.
FT TRANSIT 1 2 MITOCHONDRION (POTENTIAL).
FT CHAIN 2 576 THREONINE DEHYDRATASE.
FT BINDING 109 109 PYRIDOXAL PHOSPHATE.
FT CONFLICT 259 259 I -> T (IN REF. 1).
SQ SEQUENCE 576 AA; 63831 MW; 0801BCBD7EEDDC1F CRC64;

Query Match 22.7%; Score 395.5; DB 1; Length 576;
Best Local Similarity 32.5%; Pred. No. 7.9e-237;
Matches 104; Conservative 68; Mismatches 13; Indels 17; Gaps 10;

QY 18 NIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNATRGILPOTPEKPK 77
DB 71 SVYDVINESPISQGVGLSSRLNTNVILKREDLLPVFSFKLRGAYNMIKLA---DDSQRNQ 127
QY 78 AVYTHSSGNHGQALTYAAKLEGIPAYIVPQTAPNCKKLAIQAYGASIV-YCDPSDESRE 136
DB 128 GVIACSAGNHAQGVAAKQGLKALIVMPQNTPSIKYQNVSRGLSQVLYGNDGDEAKA 187
QY 137 KVTQRIQETEGIL-VHPNQPAPVITAGOGTTALEVLNNOVPL---VDALVVPVGGGMVAG 192
DB 188 ECAK---LAERGLTNIPFPDHPYVITAGOGTVAMEILURVETANKRIVFVPGVGGGLIAG 245
QY 193 IAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK-SSICLNTWPI 251
DB 246 IGAYLKRVAPIHIIIGVETVDAATLHNSLORNQTP-LPVVGTFADGTSYRMICEETFRV 304
QY 252 IRDLVDVFTVTEDEIKYATQLVGRMKLLIETAGVALAALVLSQHFQTVSPVEV---KNV 308
DB 305 AQQVWDEVVLNVDICAQVAKVDFEDTRISVPSGALSVAG-MKKYISTVHPHIDHTKNT 363
QY 309 CI-VLSGGNVDLTSLNNVGO 327
DB 364 YVPILSGANNFDRLRVSE 383

RESULT 11
THDH_SALTY STANDARD; PRT; 514 AA.
ID THDH_SALTY
AC P20506; O9L6S8;
DT 01-FEB-1991 (Rel. 17, Created)

```



DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 DE deaminase).  
 GN ILVA OR STM3905 OR STMD1.87.  
 OS Salmonella typhimurium.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88255870; PubMed=3290055;  
 RA Tallon B.E., Little R., Lawther R.P.;  
 RT "Analysis of the functional domains of biosynthetic threonine  
 RT deaminase by comparison of the amino acid sequences of three  
 RT wild-type alleles to the amino acid sequence of biodegradative  
 RT threonine deaminase.";  
 RL Gene 63:245-252(1988).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 [3]  
 RN SEQUENCE OF 1-10 FROM N.A.  
 RP STRAIN=LT2;  
 RX MEDLINE=89326124; PubMed=2473940;  
 RA Lopes J.M., Lawther R.P.;  
 RT "Physical identification of an internal promoter, *ilvAp*, in the  
 RT distal portion of the *ilvGMDA* operon.";  
 RL Gene 76:255-269(1989).  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHEREAS  
 CC VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M26670; AAA27150.1;  
 DR EMBL; AF233324; AAA33479.1;  
 DR EMBL; AE008882; AAF22755.1;  
 DR EMBL; M25498; AAA27151.1;  
 DR PIR; JT0278; DWEBTT.  
 DR HSP; P04968; ITDJ.  
 DR SSGene; SG10179; ilvA.  
 DR InterPro; IPR001926; B6\_enzyme\_beta.  
 DR InterPro; IPR000634; S/T\_dehydrtse.  
 DR InterPro; IPR001721; ThrDh\_C.  
 DR Pfam; PF00291; PALP; 1.  
 DR Pfam; PF00585; Thr.dehydrat.C; 2.  
 DR TIGRams; TIGR01124; ilvA\_2Cterm; 1.  
 DR PROSITE; PS00185; DEHYDRATASE\_SER\_THR; 1.  
 KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;

KW Allosteric enzyme; Complete proteome.  
 FT BINDING 62 PYRIDOXAL PHOSPHATE.  
 FT CONFLICT 71 A -> T (IN REF. 1).  
 FT CONFLICT 124 F -> L (IN REF. 1).  
 FT CONFLICT 339 A -> G (IN REF. 1).  
 FT CONFLICT 342 A -> T (IN REF. 1).  
 FT CONFLICT 351 SFL -> NEP (IN REF. 1).  
 SQ SEQUENCE 514 AA; 56253 MW; 136BC535F1F0035B CRC64;  
  
 Query Match 22.1%; Score 384; DB 1; Length 514;  
 Best Local Similarity 32.3%; Pred. No. 5.2e-22;  
 Matches 102; Conservative 69; Mismatches 127; Indels 18; Gaps 8;  
  
 QY 19 IQDSIHLPVLTSSILLNQIAGRNLFKCELFQKSGKIRGALNAIRGLPDPPEPKPA 78  
 Db 25 VYEAQVTPLOKMEKJSSRLDNVILVKREDROPVSEKRLGAYAMMAGL---TEEQKAHG 81  
 QY 79 VVTHSSGNHGQALTYAAKLEGPAYIVVPTAPNCKKLAIQAYGASIVYCDPS-DESKREK 137  
 Db 82 VITASAGNHAQGVAFSSARLGVSILVMPKATADIKVDVAVRGEGEVLLHGANFDEAKAK 141  
 QY 138 VTQRIMQETEGILVHPNQEPAVIAGOGTIALEVLNQVPLVDALVVPVGGGMVAGIAITI 197  
 Db 142 AIE-LAQOQGFVPPFDHPMVIAGOGTIALELLQDQSHLDLDRVFPVVGGLAAGVAULI 200  
 QY 198 KALKPSVKVYAAEPSNADD--CYOSKLGELTNLHPPTIADGVK-SSIGLNTWPIIRD 254  
 Db 201 KQMPQIKVIAVE---AEDSACLKALEAGHPVDLPVGLFAGGVAVKRGIDETFRUCQE 257  
 QY 255 LVDDVFVTVEDEIKYATQLVWGRMKLLIEPTAGVALAAV---LSQHFQTVSPEVKNVCIV 311  
 Db 258 YLDDIITVDSDAICAAMKDLFEDVRAVAPSGALAGMKKYIAQH---NIRGERLAHV 313  
 QY 312 LSGGNVDLTLNWNVGQ 327  
 Db 314 LSGANVNFHGLRYVSE 329  
  
 RESULT 12  
 THDH\_ARXAD  
 ID THDH\_ARXAD STANDARD; PRT; 550 AA.  
 AC 042615;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase, mitochondrial precursor (EC 4.2.1.16)  
 DE (Threonine deaminase).  
 GN ILV1.  
 OS Arxula adeninivorans.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae; Arxula.  
 OX NCBI\_TaxID=37620;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=LS3;  
 RX MEDLINE=98398049; PubMed=9730281;  
 RA Wartmann T., Roessel H., Kunze I., Bode R., Kunze G.;  
 RT "ALLV1 gene from the yeast *Arxula adeninivorans* LS3 -- a new selective  
 RT transformation marker.";  
 RL Yeast 14:1017-1025(1998).  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:36:57 ; Search time 19,5287 Seconds  
(without alignments)  
722.114 Million cell updates/sec

Title: US-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYICISFADVEKAHINR.....SSITWVQAERPASQSVSV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1735	100.0	340	1 SRR_HUMAN	Q9gzt4 homo sapien
2	1582.5	91.2	339	1 SRR_MOUSE	Q9qz77 mus musculu
3	605	34.9	326	1 YKX8_YEAST	P36007 saccharomyc
4	565	32.6	323	1 YCNE_SCHPO	O59791 schizosacch
5	471	27.1	329	1 THD2_SALTY	P11954 salmoneella
6	457	26.3	329	1 THD2_ECOLI	P05792 escherichia
7	445	25.6	332	1 Y4TJ_RHISN	P55664 rhizobium s
8	404	23.3	313	1 THD1_FASMU	Q9ckj2 pasteurellia
9	395.5	22.8	513	1 THD1_HAEIN	P46493 haemophilus
10	390.5	22.5	576	1 THDH_YEAST	P00927 saccharomyc
11	376	21.7	422	1 THD1_BACSU	P37946 bacillus su
12	375	21.6	507	1 THD1_BURCE	P53607 burkholderi
13	369.5	21.3	550	1 THDH_ARXAD	O42615 arxula aden
14	368	21.2	592	1 THD1_ARATH	Q92856 arabidopsis
15	363	20.9	595	1 THD1_LYCES	P25306 lycopersico
16	362	20.9	514	1 THD1_SALTY	P20506 salmoneella
17	357	20.6	514	1 THD1_ECOLI	P04968 escherichia
18	351.5	20.3	416	1 THD1_LACLA	Q02145 lactococcus
19	351	20.2	590	1 THD1_CICAR	Q93469 cicar ariet
20	347	20.0	415	1 THD1_BACHD	Q9kc63 bacillus ha
21	314.5	18.1	429	1 THD1_MYCTU	Q10766 mycobacteri
22	311.5	18.0	436	1 THD1_CORGL	Q04513 corynebacte
23	301.5	17.4	427	1 THD1_MYCLE	Q9x7f1 mycobacteri
24	267.5	15.4	362	1 SDHL_RAT	P09367 rattus norv
25	265.5	15.3	328	1 SDHL_HUMAN	P20132 homo sapien
26	210	12.1	359	1 THD1_SOLTU	P31212 solanum tub
27	209.5	12.1	311	1 CYSM_BACSU	O34476 bacillus su
28	203	11.7	352	1 THRC_BACSP	P09123 bacillus sp
29	201.5	11.6	354	1 THRC_BACHD	Q9k7e3 bacillus ha
30	201	11.6	307	1 CYSK_BACSU	P37887 bacillus su
31	200.5	11.6	352	1 THRC_BACSU	P04990 bacillus su
32	200	11.5	382	1 THRC_SYNY3	P74193 synechocyst
33	198.5	11.4	312	1 CYSK_SYNY3	P73410 synechocyst

RESULT 1  
SRR\_HUMAN  
ID SRR\_HUMAN STANDARD; PRT; 340 AA.  
AC Q9GZT4;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serine racemase (EC 5.1.1.1-).  
GN SRR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20510003; PubMed=11054547;  
RA DeMiranda J., Santoro A., Engelender S., Wolosker H.;  
RT "Human serine racemase: molecular cloning, genomic organization and  
functional expression.";  
RL Gene 256:183-188(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
Ninomiya K., Iwayanagi T.;  
RL "NEO human cDNA sequencing project.";  
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.  
CC -!- TISSUE SPECIFICITY: Brain. Expressed at high levels in hippocampus  
and corpus callosum, intermediate levels in substantia nigra and  
caudate, and low levels in amygdala, thalamus, and subthalamic  
nuclei.  
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: AF169974; AG27081.1; - Oct 27, 2000  
EMBL: AK023169; BAB14442.1;  
Genew; HGNC:14398; SRR.  
MIM; 606477;  
InterPro: IPR001926; B6\_enzyme\_beta.  
InterPro: IPR000634; S/T\_dehydrtse.  
Pfam: PF00291; PALP; 1.  
PROSITE: P500165; DEHYDRATASE\_SER\_THR; 1.  
Isomerase; Pyridoxal phosphate.

34 196.5 11.3 338 1 SDHL\_YEAST  
35 191 11.0 310 1 CYSK\_MYCTU  
36 189 10.9 307 1 CYSK\_FLASP  
37 187 10.8 360 1 THRC\_MYCLE  
38 187 10.8 550 1 CBS\_HUMAN  
39 184.5 10.6 310 1 CYSK\_MYCLE  
40 181 10.4 360 1 THRC\_MYCTU  
41 174.5 10.1 323 1 CYSM\_MYCTU  
42 173.5 10.0 336 1 Y4XP\_RHISN  
43 172 9.9 560 1 CBS\_RAT  
44 171.5 9.9 405 1 THRC\_METJA  
45 171 9.9 299 1 CYSM\_CAMJE

ALIGNMENTS

P17324 saccharomyc  
P95230 mycobacteri  
Q59447 flavobacter  
P45837 mycobacteri  
P35320 homo sapien  
O32978 mycobacteri  
Q10610 mycobacteri  
Q10624 mycobacteri  
P55708 rhizobium s  
P32232 rattus norv  
Q58860 methanococc  
P71128 campylobact

```
FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 340 AA; 36566 MW; 873342C62D5D7B9D CRC64;

Query Match 100.0%; Score 1735; DB 1; Length 340;
Best Local Similarity 100.0%; Pred. No. 1e-125;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSITLNLGTGRNLFKCELFQKTSFQKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSITLNLGTGRNLFKCELFQKTSFQKIRGA 60
QY 61 LNAVRLSPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIOA 120
DB 61 LNAVRLSPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIOA 120
QY 121 YGASIVYCPDSRENRVAKRVTETEGIMVHPNQEPAVIAGQGTIALEVLNQVPLVDAL 180
DB 121 YGASIVYCPDSRENRVAKRVTETEGIMVHPNQEPAVIAGQGTIALEVLNQVPLVDAL 180
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPETIADGV 240
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPETIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAVALSQHFQT 300
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAVALSQHFQT 300
QY 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPRASYOSVS 340
DB 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPRASYOSVS 340

RESULT 2
SRR_MOUSE STANDARD; PRT; 339 AA.
ID SRR_MOUSE
AC Q9QZK7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Serine racemase (EC 5.1.1.-)
GN SRR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=20027561; PubMed=10557334;
RA Wolosker H., Blackshaw S., Snyder S.H.;
RT "Serine racemase: a glial enzyme synthesizing D-serine to regulate
RT glutamate-N-methyl-D-aspartate neurotransmission.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13409-13414(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the synthesis of D-serine from L-serine.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF148321; AAF08701.1; -
CC EMBL: BC01164; AAH1164.1; -
CC HSSP: P04968; ITD.
```

```
DR MGD; MGI:1351636; Srr.
DR InterPro: IPR001926; B6_enzyme_beta.
DR InterPro: IPR000634; S/T_dehydrtse.
DR Pfam: PF00291; PALP; 1.
DR PROSITE: P500165; DEHYDRATASE_SER_THR; 1.
KW Isomerase; Pyridoxal phosphate.
FT BINDING 56 56 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 339 AA; 36359 MW; B9AE9A9336358728 CRC64;

Query Match 91.2%; Score 1582.5; DB 1; Length 339;
Best Local Similarity 89.7%; Pred. No. 4.9e-114;
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;

QY 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSITLNLGTGRNLFKCELFQKTSFQKIRGA 60
DB 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSITLNLGTGRNLFKCELFQKTSFQKIRGA 60
QY 61 LNAVRLSPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIOA 120
DB 61 LNAVRLSPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPPQAPDCCKLAIOA 120
QY 121 YGASIVYCPDSRENRVAKRVTETEGIMVHPNQEPAVIAGQGTIALEVLNQVPLVDAL 180
DB 121 YGASIVYCPDSRENRVAKRVTETEGIMVHPNQEPAVIAGQGTIALEVLNQVPLVDAL 180
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPETIADGV 240
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPETIADGV 240
QY 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAVALSQHFQT 300
DB 241 KSSIGLNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAVALSQHFQT 300
QY 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPRASYOSVS 340
DB 301 VSPEVKNICIVLSGGNVDLTSSITWVKQAEPRASYOSVS 339

RESULT 3
YKV8_YEAST STANDARD; PRT; 326 AA.
ID YKV8_YEAST
AC P36007;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 34.9 kDa protein in COS9-JEN1 intergenic region.
GN YKL218C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=95028164; PubMed=7941750;
RA Tzermia M., Horaitis O., Alexandraki D.;
RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI
RT identified the known loci URA1, SAC1 and TRP3, and revealed 6 new
RT open reading frames including homologues to the threonine
RT dehydratases, membrane transporters, hydantoinases and the
RT phospholipase A2-activating protein.";
RL Yeast 10:663-679(1994).
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```

DR EMBL: X75951; CAA53555.1; -  
 DR EMBL: Z28218; CAA82063.1; -  
 DR PIR: S38061; S38061.  
 DR PIR: S44320; S44320.  
 DR HSP: P04968; LTDJ.  
 DR SGD: S0001701; YKL218C.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR Pfam: PF00291; PALP; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).  
 FT BINDING 53 53  
 SQ SEQUENCE 326 AA; 34899 MW; 21CFEFC9AB4431 CRC64;

Query Match 34.9%; Score 605; DB 1; Length 326;  
 Best Local Similarity 41.0%; Pred. No. 3.2e-39;  
 Matches 133; Conservative 64; Mismatches 119; Indels 8; Gaps 4;

QY 8 SPADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67  
 DB 5 TYGDVLDASNRKKEYVKNKTPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 64  
 QY 68 VPDALERKPKAVVTHSSNGHQAALYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 127  
 DB 65 ----SDEKRSKGVIAFSSNGHQAALYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 121  
 QY 128 CEPDSRENVAKRVTEETEGIMVHPNOBPVIAAGGTIALEVNOVPLVDALVVPVGGG 187  
 DB 122 YNRYTEDREIQGLAAEHGFALIPYDHPDVIAGQTSKELLEBEVGGLDGLFVPLGGG 181  
 QY 188 GMLAGIATVTKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGVKSS-IGL 246  
 DB 182 GLLSGSALAARSLSPCKIFGVEPEAGNDGQOSFRSGSIV-HINTPKTIADGAQTQHLGE 240  
 QY 247 NTWPIIRLDVDDIFVTDEIKATQLVWERKLLIEPTAGVGVAAVLSQHFQTVSPVK 306  
 DB 241 YFAIIRENVDDILTVSQELVKMHFLAERMKVVEPTACLGFAAGALKKEELVG---K 297  
 QY 307 NTCIVLSGGNVLTSSITWVQAE 330  
 DB 298 KVGIIILSGGNDMKRYATLISGKE 321

## RESULT 4

YCNESCHPO  
 ID YCNE SCHPO STANDARD; PRT: 323 AA.  
 AC O59791;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C320.14 in chromosome III.  
 GN SPCC320.14 OR SPCC330.15C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 ON NCBI\_TaxID=4896;  
 RX STRAIN=972;  
 RC SEQUENCE FROM N.A.  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880 (2002).  
 CC - COPACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AL022245; CAA18316.1; -  
 DR EMBL: AL031603; CAA20920.1; -  
 DR HSP: P04968; LTDJ.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR Pfam: PF00291; PALP; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; FALSE NEG.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate. (BY SIMILARITY).  
 FT BINDING 57 57  
 SQ SEQUENCE 323 AA; 35048 MW; 21187E9A69FA5348 CRC64;

Query Match 32.6%; Score 565; DB 1; Length 323;  
 Best Local Similarity 40.3%; Pred. No. 3.7e-36;  
 Matches 127; Conservative 67; Mismatches 109; Indels 12; Gaps 6;

QY 8 SPADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67  
 DB 9 TYDDVASASERIKKFKANKTPVLTSSITVNFKEFVAEVEFKCFKMGAFKRGALNALSQL 68  
 QY 68 VPDALERKPKAVVTHSSNGHQAALYAAKLEGIPAYIVVPTAPDCKKLAIOAGYASIVY 127  
 DB 69 --NEAQRK-AGVLTSSNGHQAALYAAKLEGIPAKIIMPLDAPEAKVAAVTKYGCQVIM 125  
 QY 128 CEPDSRENVAKRVTEETEGIMVHPNOBPVIAAGGTIALEVNOVPLVDALVVPVGGG 187  
 DB 126 YDRYKDDREKMAKEISEREGLTIIPYDHPVLAGGTAAKELFEVGLDALFVCLGGG 185  
 QY 188 GMLAGIATVTKALKPSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGVKSS-IGL 246  
 DB 186 GLLSGSALAARHFAFNPCEYVEPEAGNDGQOSFRKGSIV-HIDTPTIADGAQTQHLGN 244  
 QY 247 NTWPIIRLDVDDIFVTDEIKATQLVWERKLLIEPTAGVGVAAVLSQHFQTVSPVK 306  
 DB 245 YTFSIKKEKVDILTVSDDELIDCLFYAARMKIVVEPTGCLSLFAAA-----RAMKEK 299  
 QY 307 N-ICIVLSGGNVDL 319  
 DB 300 NKRIIGIISGGNVDI 314

RESULT 5  
 THD2\_SALTY  
 ID THD2\_SALTY STANDARD; PRT: 329 AA.  
 AC P11954;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).

GN TDCB OR STM3244.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Letreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE OF 1-24;  
RX MEDLINE=83023208; PubMed=6751404;  
RA Kim S.S., Datta P.;  
RT "Chemical characterization of biodegradative threonine dehydratases  
RT from two enteric bacteria.";  
RL Biochim. Biophys. Acta 706:27-35(1982).  
CC -|- CATALYTIC ACTIVITY: L-threonine + H(2)O -> 2-oxobutanoate + NH(3) +  
CC H(2)O.  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -|- ENZYME REGULATION: EACH PROTEIN MOLECULE CAN BIND UP TO FOUR  
CC MOLECULES OF AMP, WHICH ACTS AS AN ALLOSTERIC ACTIVATOR TO THE  
CC ENZYME. THE ENZYME IS ALSO INHIBITED BY ALPHA-KETO ACIDS AND OTHER  
CC CATABOLITES. IT IS ALLOSTERICALLY INHIBITED BY ISOLEUCINE  
CC AND ALLOSTERICALLY ACTIVATED BY VALINE.  
CC -|- PATHWAY: Threonine catabolism.  
CC -|- SUBUNIT: HOMOTETRAMER.  
CC -|- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE008849; AAL22117.1; --  
CC PIR: B22317; B22317.  
CC StyGene: SGI0390; tdcB.  
CC InterPro: IPR001926; B6\_enzyme\_beta.  
CC InterPro: IPR000634; S/T\_dehydratse.  
CC Pfam: PF00291; PALP; 1.  
CC PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
KW Lyase; Pyridoxal phosphate; Allosteric enzyme; Complete proteome.  
FT BINDING 58 58 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT CONFLICT 24 24 K -> I (IN REF. 2).  
SQ SEQUENCE 329 AA; 35141 MW; C1C619B021DE817C CRC64;  
Query Match 27.1%; Score 471; DB 1; Length 329;  
Best Local Similarity 36.6%; Pred. No. 5.9e-29;  
Matches 123; Conservative 64; Mismatches 121; Indels 28; Gaps 10;  
QY 7 ISFADVEKAHNIIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTGSKIRGALNAVR 66  
DB 9 VAIEDILEAKRLAGKIYKTPRMSNYFSERCKGEIFLKFENMQRTGSPKIRGAENKLSS 68  
QY 67 LVPDALEKPKAVVTHSSNHQALTYAAKLGEGIPAYIVVQTPADPCKKLATQAGSIV 126  
DB 69 LFE---AERKKGWVACSNAGNHAQGVSLSCAMLGIDGKVVMPKGPASKVAATCDYSAEV 125  
QY 127 YCEPDESRENKAVRTE--ETEG-IMVHPNPEVPIAGQGTIALEVLNOVPLVDALVVP 183  
DB 126 L--HGDNFNDTAK-VSEIVETEGRIFFIPYDDPKVPIAGQGTIGLEIMEDLYDENVIVP 182  
QY 184 VGGGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLPNLYPPETIADGVK-S 242

Db 183 IGGGGLIAGIAIAIKSINPTIKVIGQAEVNHGMAASYVTGETTH-RTTGTADCCDVS 241  
QY 243 STGLNTWPIIRDIVDDIFTVTEDEIKCATQLVWRMKLLIETAGVGVAIVS----- 295  
DB 242 RPNLTYYEIVRELVDDIVLVSEDEIRNSMIALIQRNKVITEGAGALACAAALLSGKLDHSI 301  
QY 296 QHFQTVSPVKNICIVLGGNVDLT--SSITVWKQA 329  
DB 302 QNRKTVS-----IISGNNIDLSRVSKVQITGLVDA 329  
RESULT 6  
THD2\_ECOLI  
ID THD2\_ECOLI STANDARD; PRT; 329 AA.  
AC P05792;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Threonine dehydratase catabolic (EC 4.2.1.16) (Threonine deaminase).  
GN TDCB OR B3117 OR Z4469 OR ECS3997.  
OS Escherichia coli, and  
OS Escherichia coli, O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=87092415; PubMed=3540965;  
RA Datta P., Goss T.J., Omnas J.R., Patil R.V.;  
RT "Covalent structure of biodegradative threonine dehydratase of  
RT Escherichia coli: homology with other dehydratases.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:393-397(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=89282418; PubMed=2660107;  
RA Schweizer H., Datta P.;  
RT "The complete nucleotide sequence of the tdc region of Escherichia  
RT coli.";  
RL Nucleic Acids Res. 17:3994-3994(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli





```
QY 238 DGKSSIGL-NTW--PIIRDVLVDIFTVTDEIKCATQLVWERMKLLIEPTAGVGVAVL 294
Db 234 DSLGGGIGNANAWTFQMCALLDDVVLNVEGEIAAGIRHAYEHERQILEGAGAVGIAALL 293
QY 295 SQHFQTVSEVKNICIVLSGGNDL 319
Db 294 S---GRVAAARGSGVGVLSGQNDM 315

RESULT 8
THDL_PASMU STANDARD; PRT; 513 AA.
AC Q9CKJ2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
deaminase).
GN ILVA OR PM1624.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
threonine in a two-step reaction. The first step is a dehydration
of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +
H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006199; AAK03708.1; -
CC HSSP; P04968; ITDJ.
CC InterPro; IPR001926; B6_enzyme_beta.
CC InterPro; IPR000634; S/T_dehydrtse.
CC InterPro; IPR001721; ThrDh_C.
CC Pfam; PF00291; PALP; 1.
CC TIGRFAMs; TIGR01124; ilva_2c term; 1.
CC PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW Complete proteome.
FT BINDING 61 PYRIDOXAL PHOSPHATE.
SQ SEQUENCE 513 AA; 56288 MW; 3E427ADC54E5FC7C CRC64;

Query Match 23.3%; Score 404; DB 1; Length 513;
Best Local Similarity 32.6%; Pred. No. 1.4e-23;
Matches 98; Conservative 65; Mismatches 132; Indels 6; Gaps 4;
QY 19 IRDSIHLTPVLSSILNQITGNRLNFKBELFKTGSFKIRGNVRSVLPALERKPKA 78
Db 24 YVDVAQVTPLODMAKLSERLGNKVFTRKDRQPVHSFKLRGAYAMTAGL---SAEQKASG 80
QY 79 VVTHSSGNHGOALTAAKLEGIPAYIVVPQAPDCCKKLAIOYAGASIVYCEPSDESRENV 138
Db 81 VTAASAGNQAQVALSAKHLGRLALIVMPONTPSIKVDVAVRGFGVGLLHGANFDEAKAK 140
```

```
QY 139 AKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNVPLVDALVVPYGGGMLAGIAITVK 198
Db 141 ATELAESKNMTFIPPDHPAVIAGOGSIAMELLQONSQIDRFVPVGGGGLAAGIAVLK 200
QY 199 ALKPSVYVYAABPSNADDCYQSKLGLMPNLYPPETIADGVK-SSIGLNTWPIIRDVLVD 257
Db 201 QLMPEIKVIGVESKDSACLYRALKAGKPI-DLDRVGLFADGVAVKRGIDETFRVCCQYID 259
QY 258 DIFTYTEDEIKCATQLVWERMKLLIEPTAGVGVAVLSSOHFQTVSEVKNICIVLSGGNV 317
Db 260 DVVLVDGDEICAAVKDIFENRAIEPSPGALSGLAG-LKKYVKEHNTQGETLVNLSGANL 318
QY 318 D 318
Db 319 N 319

RESULT 9
THDL_HAEIN STANDARD; PRT; 513 AA.
AC P46493;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine
deaminase).
GN ILVA OR HI0738.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RN Science 269:496-512(1995).
RL [2]
RP IDENTIFICATION.
RA Koonin E.V., Rudd K.E.;
RL Submitted (SEP-1995) to the SWISS-PROT data bank.
CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from
threonine in a two-step reaction. The first step is a dehydration
of threonine, followed by rehydration and liberation of ammonia.
CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +
H(2)O.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Isoleucine biosynthesis; first step.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32757; AAC22398.1; -
CC HSSP; P04968; ITDJ.
CC TIGR; HI0738.1; -
```



DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 DE deaminase).  
 GN ILVA.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=1423;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Arnpriester J.M. Jr., Fink P.S.;  
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -!- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M58606; AAA22549.1; -  
 DR EMBL; L77246; AAA96639.1; -  
 DR EMBL; Z99115; CAB14095.1; -  
 DR HSP; P04968; ITD; -  
 DR Subtilisin; BG10673; ILVA.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydrtse.  
 DR InterPro: IPR001721; ThrDh\_C.  
 DR Pfam; PF00291; PALP; 1.  
 DR Pfam; PF00585; Thr\_dehydrat\_C; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;  
 KW Complete proteome.  
 FT BINDING 60 60 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 15 15 H -> D (IN REF. 2).  
 FT CONFLICT 139 139 V -> A (IN REF. 2).  
 SQ SEQUENCE 422 AA; 46701 MW; 3027A5ED87084140 CRC64;  
 Query Match 21.7%; Score 376; DB 1; Length 422;  
 Best Local Similarity 31.1%; Pred. No. 1 5e-21;  
 Matches 99; Conservative 64; Mismatches 143; Indels 12; Gaps 7;  
 QY 7 ISPADVEKAHINIRDSIHLPVLTSSILNLTGRNLFFKCELFQKTSFKIRGALNAVR.66  
 Db 11 IQVKHILKAHQNVKQVVIHTPLQRNDRLSERYECNIVLKREDLQVVRFSFKRGAYHKMQ 70  
 QY 67 LVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPDCKKLAIOAGSIV 136  
 Db 71 L--SSQETENGVCVCSAGNHAQGVAFSKHLGIHKIFMPSTTPKQVSVQVLFQKGF 127  
 QY 127 YCPSPDESRENVAKRVT--ETEG-IMVHPNQPAVITAGOGTIALEVLNOVPL-VDAVV 182  
 Db 128 DILTGDTDDVYKSAECCESRTFIHFPDPPVWAGOGTLAVEILNDIDTEPHLFA 187  
 QY 183 PVGGGMLAGIATVYKALPKSVKYAAEPSNADCCYOSKLGKLMPLNLYPETIADGVK - 241  
 Db 188 SVGGGGLSGVGYLKNVSPDTKVIAPVAPAGAASYFESNKGHWV-TLDKIDKFDVGA 246  
 QY 242 SSTGLNTPIIRDVLDDITVTEDEIKCATQLVYERMKLLIETAGVGVAAVLSDHFTV 301  
 Db 247 KKGTEFTTLETVDVDDILLVPEGKVTSTILEYNECAVVAEPAGALSVAALDYKDI - 305  
 QY 302 SPEVKNICIVLSGGNVDL 319  
 Db 306 --KGKVVVCVSGGNNDI 321  
 RESULT 12  
 THD1\_BURCE STANDARD; PRT; 507 AA.  
 AC P53607;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine dehydratase biosynthetic (EC 4.2.1.16) (Threonine  
 DE deaminase).  
 GN ILVA.  
 OS *Burkholderia cepacia* (Pseudomonas cepacia).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC *Burkholderia*.  
 RN NCBI\_TaxID=292;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=17616;  
 RA Bartell J.B., Lessie T.G.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -!- CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanate + NH(3) +  
 CC H(2)O.  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.





Job time : 20.5287 secs

```
OC Asteridae: euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RX STRAIN=CV, TINY TIM LA154;
RA MEDLINE=91187855; PubMed=2011578;
RT Samach A., Hareven D., Gutfinger T., Ken-Dror S., Lifschitz E.;
RT "Biosynthetic threonine deaminase gene of tomato: isolation,
RT structure, and upregulation in floral organs.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2678-2682(1991).
CC -! CATALYTIC ACTIVITY: L-threonine + H(2)O = 2-oxobutanoate + NH(3) +
CC H(2)O.
CC -! COFACTOR: PYRIDOXAL PHOSPHATE.
CC -! ENZYME REGULATION: ALLOSTERICALLY INHIBITED BY ISOLEUCINE.
CC -! PATHWAY: Isoleucine biosynthesis; first step.
CC -! SUBUNIT: HOMOTETRAMER.
CC -! SUBCELLULAR LOCATION: Chloroplast.
CC -! MISCELLANEOUS: EXPRESSION IN MATURE FLOWERS IS INCREASED OVER
CC 500-FOLD.
CC -! SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61914; AAA34171.1; -.
DR EMBL: M61915; AAA68097.1; -.
DR PIR: A38628; A38628.
DR HSSP: P04968; 1TDJ.
DR InterPro: IPR001926; B6 enzyme beta.
DR InterPro: IPR000634; S/T dehydratse.
DR InterPro: IPR001721; ThrDh_C.
DR Pfam: PF00291; PALP; 1.
DR Pfam: PF00585; Thr_dehydrat_C; 2.
DR TIGRFAMs: TIGR01124; ilvA_2Cterm; 1.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
DR Isoleucine biosynthesis; Lyase; Chloroplast; Pyridoxal phosphate;
KW Allosteric enzyme; Transit peptide.
FT TRANSIT 1 51 CHLOROPLAST.
FT CHAIN 52 595 THREONINE DEHYDRATASE BIOSYNTHETIC.
FT BINDING 143 143 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 595 AA; 64937 MW; AC430BB5DD9F0348 CRC64;

Query Match 20.9%; Score 363; DB 1; Length 595;
Best Local Similarity 32.1%; Pred. No. 2.3e-20;
Matches 96; Conservative 62; Mismatches 129; Indels 12; Gaps 8;

QY 26 TPVLTSSILNLTGRNLFKCELFQKSGFKIRGALNAVRLPYDALKERPKAVVTHSSG 85
Db 113 SPLELAEKLSDRIGVNFYIKREDKQRFKFLRGAYNMMSNLSRELD---KGVITASAG 169

QY 86 NHGQALTYAAKLEGIPAYIVPOTAPDCKKLAIQAYCASIV-YCEPSDESRENVAKRUTE 144
Db 170 NHAQGVALAGQRLNCVAKIVMPTTTPQIKIDAVRALGGDVLYGKTFDEAQTHALE--LS 227

QY 145 ETEGI-MVHPNQBPVAGIAGTALVNOVPLVDALVVPVGGGMLAGIAITVTKPKPS 203
Db 228 EKDLGYIIPFPDGPVIGKGGTIGTEINRQLKDIHAVFIPVGGGLIAGVATFFKQIAPN 287

QY 204 VKVYAAEPSNADCYOSKLGKMLPNLYPETIADGVK-SSICLNTWPIIRDLVDDIFTV 262
Db 288 TKIIGVEPYGAASMTLSLHEGHRV-KLSNVDTFADGVAVALVGEXTFAKQCELIDGMVLV 346

QY 263 TEDEIKCATQLVWERMKLLIETAGVGAVALISO-HFQTVSPVKNICIVLSGGNVDLT 320
Db 347 ANDGISAAIKDYDEGRNILETSGAVAIAGAAAYCEFYKIKNE--NIVATASGANMDFS 403
```

Search completed: June 24, 2003, 06:19:31

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 02:01:55 ; Search time 57 Seconds  
(without alignments)  
5477.139 Million cell updates/sec

Title: US-09-889-609B-1  
Perfect score: 1018  
Sequence: 1 atgtgtctcagtgactgcgt.....ttaccagaagcgtctgttttaa 1018

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	807.8	79.4	1020	4	US-09-789-300A-3
2	807.8	79.4	1770	4	US-09-789-300A-1
3	62.2	6.1	2299	4	US-09-153-599A-1
4	60.6	6.0	1545	2	US-08-628-039-1
5	60.6	6.0	1545	2	US-08-628-039-5
6	60.6	6.0	1545	2	US-08-628-039-7
7	60.6	6.0	1545	2	US-08-628-039-8
8	60.6	6.0	1545	2	US-08-673-388-1
9	60.6	6.0	1545	2	US-08-673-388-5
10	60.6	6.0	1545	2	US-08-673-388-7
11	60.6	6.0	1545	2	US-08-673-388-8
12	60.6	6.0	1545	2	US-08-614-877-1
13	60.6	6.0	1545	2	US-08-614-877-5
14	60.6	6.0	1545	2	US-08-614-877-7
15	60.6	6.0	1545	2	US-08-614-877-8
16	60.6	6.0	1545	3	US-08-912-205-1
17	60.6	6.0	1545	3	US-08-912-205-5
18	60.6	6.0	1545	3	US-08-912-205-7
19	60.6	6.0	1545	3	US-08-912-205-8
20	60.6	6.0	1545	4	US-09-440-400-1
21	60.6	6.0	1545	4	US-09-440-400-5
22	60.6	6.0	1545	4	US-09-440-400-7
23	60.6	6.0	1545	4	US-09-440-400-8
24	52.8	5.2	12720	1	US-08-403-866-11
25	45.6	4.6	7218	1	US-08-232-463-14
26	38.8	3.8	1485	4	US-09-088-435-2
27	34.4	3.4	1275	4	US-09-134-001C-1039

c	28	34	3.3	3655	3	US-08-878-474-6	Sequence 6, Appli
c	29	33.8	3.3	1274	4	US-09-227-357-75	Sequence 75, Appl
c	30	33.4	3.3	2698	4	US-08-936-165A-207	Sequence 207, App
c	31	32.8	3.2	1956	4	US-08-559-896B-1	Sequence 1, Appli
c	32	32.6	3.2	3006	4	US-09-449-285A-1	Sequence 1, Appli
c	33	31.8	3.1	3998	4	US-09-066-046-5	Sequence 5, Appli
c	34	31.2	3.1	1419	1	US-08-103-739B-1	Sequence 1, Appli
c	35	31.2	3.1	1419	1	US-08-474-404-1	Sequence 1, Appli
c	36	31.2	3.1	1419	2	US-08-485-845-1	Sequence 1, Appli
c	37	31.2	3.1	1419	2	US-08-482-714-1	Sequence 1, Appli
c	38	31.2	3.1	1419	4	US-09-211-416-1	Sequence 1, Appli
c	39	31.2	3.1	1419	4	US-09-059-958-1	Sequence 1, Appli
c	40	31.2	3.1	1419	4	US-09-221-017B-1092	Sequence 1092, Ap
c	41	31	3.0	2802	4	US-09-484-970B-25	Sequence 25, Appli
c	42	31	3.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	43	31	3.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	44	30.8	3.0	382	4	US-09-118-554-59	Sequence 59, Appli
c	45	30.8	3.0	382	4	US-09-118-627-59	Sequence 59, Appli

ALIGNMENTS

RESULT 1  
US-09-789-300A-3  
; Sequence 3, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; FILE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq For Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-789-300A-3

Query Match	79.4%	Score	807.8;	DB	4;	Length	1020;
Best Local Similarity	87.7%	Pred	No. 6.8e-267;				
Matches	894;	Conservative	0;	Mismatches	122;	Indels	3;
Gaps	1;						
Qy	1	ATGTGTCTCAGTACTGCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA	60				
Db	1	ATGTGTCTCAGTACTGCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA	60				
Qy	61	GACTCTATCCACCTCACCCTCAGTGTACACAGCTCCATTTTGAATCAATACGAGGGGC	120				
Db	61	GATTCATCCACCTCACCCTCAGTGTACACAGCTCCATTTTGAATCAATACGAGGGGC	120				
Qy	121	AATCTTTTCTTCAAAATGTGAGCTCTTCAGAAAACTGGTCTTTTAAAGATTCGAGGTGC	180				
Db	121	AATCTTTTCTTCAAAATGTGAGCTCTTCAGAAAACTGGTCTTTTAAAGATTCGAGGTGC	180				
Qy	181	CTTAATCCATCAGAGGCTTAATTCCTGACAGCCGAGAGAGAGCCCAAGCCGTAGT	240				
Db	181	CTTAATCCATCAGAGGCTTAATTCCTGACAGCCGAGAGAGAGAGCCCAAGCCGTAGT	240				
Qy	241	ACTCACAGCAGGGAACCATGGCCAAAGCTCTACCTATGCTGCTAACTGGAAGGAAT	300				
Db	241	ACTCACAGCAGGGAACCATGGCCAAAGCTCTACCTATGCTGCTGCAAAATGGAAGGAAT	300				
Qy	301	CCTGCTTACATTTGGTTTCCCAACAGCTCCCAACTGCAAGAACTGCAATCAAGCC	360				
Db	301	CCTGCTTACATTTGGTTTCCCAACAGCTCCCAACTGCAAGAACTGCAATCAAGCC	360				

```
QY 361 TATGGAGCATCGATAGTACTGTGACCCAAAGTCAGAGTCCAGAGAAAGGTCACTCAA 420
Db 361 TACGGAGCGTCAATTCTATCTGTGAACCTAGTAGTGATGATCCAGAGAAATGTGCAAAA 420
QY 421 AGAATTATCGAAGAAACAGAGGCATCTTGGTCCATCCCAACAGGAGGCTGCGAGTGATA 480
Db 421 AGAGTTTACAGAGAAACAGAGGCATCATGGTACATCCCAACAGGAGGCTGCGAGTGATA 480
QY 481 GCTGGACAAGGAACAATTCCTCGGAAGTGGCTGCAACCCAGGTTCCCTTGGTAGATGACATG 540
Db 481 GCTGGACAAGGAACAATTCCTCGGAAGTGGCTGCAACCCAGGTTCCCTTGGTAGATGACATG 540
QY 541 GTGGTACCAGTAGGAGGAGGAGGAATGGTGGTGGGAATACCAATTAAGGCCCTG 600
Db 541 GTGGTACCAGTAGGAGGAGGAGGAATGGTGGTGGGAATACCAATTAAGGCCCTG 600
QY 601 AAACCTAGTGTGAGGTATACGCTGTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 660
Db 601 AAACCTAGTGTGAGGTATACGCTGTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 660
QY 661 AAACCTAGTGTGAGGTATACGCTGTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 720
Db 661 AAACCTAGTGTGAGGTATACGCTGTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 720
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAGAGAGACCTTGTGGATGATGCTTTC 780
Db 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAGAGAGACCTTGTGGATGATGCTTTC 780
QY 781 ACTGTCACCGAAGATCAATCAAGTATGCAACCCAGCTGGTGGGGGAGAGATGAACATG 840
Db 781 ACTGTCACCGAAGATCAATCAAGTATGCAACCCAGCTGGTGGGGGAGAGATGAACATG 840
QY 841 CTCATTGAGCGCACTGCTGGCTGGGCACTGGCTGCAAGTCTGCTCAGCATTTCCAAACA 900
Db 841 CTCATTGAGCGCACTGCTGGCTGGGCACTGGCTGCAAGTCTGCTCAGCATTTCCAAACA 900
QY 901 GTCTCTCCAGAAGTAAAGAACCTGCTGATTTACTCAGTGGGGGAATAGAACCTAA-- 958
Db 901 GTTTCCTCCAGAAGTAAAGAACCTTGTATTGCTCAGTGGTGGGAATAGAACCTAAACC 960
QY 959 -CCTCCCTGAACCTGGTGGGGAGGCTGAACGCCAGCTCCCTTACCAGAGGCTGCTGTTT 1016
Db 961 TCTCCATAAATTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTATCAGTCTGTTCTGT 1019
```

## RESULT 2

```
US-09-789-300A-1
; Sequence 1, Application US/09789300A
; Patent No. 6458576
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 2406, A No. 6458576el Human Pyridoxal-Phosphate
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore
; FILE REFERENCE: 35800/208926
; CURRENT APPLICATION NUMBER: US/09/789,300A
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: US 60/183,208
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (69)...(1091)
US-09-789-300A-1
```

Query Match 79.4%; Score 807.8; DB 4; Length 1770;

Best Local Similarity 87.7%; Pred. No. 9.8e-267;

Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

```
QY 1 ATGTGTGCTCAGTACTGCAATCTCCTTTGCTGTGTTGAAAGAGCTCATATCAACATTCAA 60
Db 69 ATGTGTGCTCAGTACTGCAATCTCCTTTGCTGTGTTGAAAGAGCTCATATCAACATTCGA 128
QY 61 GACTCTATCACCCTACCCAGTGTCTAACAAGCTCCATTTGAATCAATAGCAGGCGC 120
Db 129 GATTCTATCCACTCACACCAGTGTCTAACAAGCTCCATTTGAATCAATAGCAGGCGC 188
QY 121 AATCTTTTCTTCAAAATGTGAGCTTTCCAGAAACTGGGTCTTTTAAAGATTCAGAGTGCC 180
Db 189 AATCTTTTCTTCAAAATGTGAGCTTTCCAGAAACTGGGTCTTTTAAAGATTCAGAGTGCC 248
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 240
Db 249 CTCATTCGCCGTGAGAAGCTTGGTTCTGTGCTTTAGAAAGGAGGCCGAAAGCTGTGTT 308
QY 241 ACTCAGCAGGAGGAAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGAAGAAAT 300
Db 309 ACTCAGCAGTGGAAACCATGCGCAGGCTCTCACCTATGCTGCTGCAATTTGGAAGAAAT 368
QY 301 CTGTCTTACATTTGTGGTTCCTCCAAACAGCTCCCACTGCAAGAACTGGCAATTCAGGCC 360
Db 369 CTGTCTTATATTGTGGTTCCTCCAGACAGCTCCAGAGCTGTAAGAACTTGAATACAAGCC 428
QY 361 TATGGAGCATCGATAGTACTGTGACCCCAAGTGCAGAGTCCAGAGAAAGGTCACTCAA 420
Db 429 TACGGAGCTCAATTTGTACTGTGAACCTAGTGTGAGTCCAGAGAAATGTGTCAGAAA 488
QY 421 AGAATTATCAAGAAACAGAGGCACTTTGGTCCATCCCAACAGGAGGCTGCGAGTGATA 480
Db 489 ACAGTTTACAGAGAAACAGAGGCACTATGGTACATCCCAACAGGAGGCTGCGAGTGATA 548
QY 481 GCTGACAGGAGCAATTTCCCTGGAGTGTGGAAGTGTGGAATAGCCATTAACAATTAAGGCCCTG 540
Db 549 GCTGACAGGAGCAATTTCCCTGGAGTGTGGAAGTGTGGAATAGCCATTAACAATTAAGGCCCTG 608
QY 541 GTGTGTACCATGAGGAGGAGGAGGATGGTGTGGAATAGCCATTAACAATTAAGGCCCTG 600
Db 609 GTGTGTACCATGAGGAGGAGGAGGATGGTGTGGAATAGCCATTAACAATTAAGGCCCTG 668
QY 601 AAACCTAGTGTGAGGATATACGCTGCTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 660
Db 669 AAACCTAGTGTGAGGATATACGCTGCTGAGCCCTCGAATGCAAGTACGCTGCTACCAAGTCT 728
QY 661 AAACCTAGTGTGAGGATATACGCTGCTGAGCCCTCGAATGCAAGTACGCTGCTGCTGCT 720
Db 729 AAGCTGAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG 788
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATAGAGAGACCTTGTGGATGATGCTTTC 780
Db 789 AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATAGAGAGACCTTGTGGATGATGCTTTC 848
QY 781 ACTGTACCGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGGAGAGATGAACATG 840
Db 849 ACTGTACCGAAGATGAATTAAGTGTGCAACCCAGCTGGTGTGGGGGAGAGATGAACATG 908
QY 841 CTCATTGAGCCAGCTGCTGGCCGTGGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 909 CTCATTGAGCCAGCTGCTGGCCGTGGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 968
QY 901 GTCTCTCCAGAAGTAAAGAACCTGCTGCAATTTACTCAGTGGGGGGAATAGAACCTAA-- 958
Db 969 GTTTCCTCCAGAAGTAAAGAACCTTGTATTGCTCAGTGGTGGGAATAGAACCTTAACC 1028
QY 959 -CCTCCCTGAACCTGGTGGGGAGGCTGAACGCCAGCTCCCTTACCAGAGGCTGCTGTTT 1016
Db 1029 TCTCCATAAATTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTATCAGTCTGTTCTGT 1087
```

## RESULT 3

US-09-153-599A-1/c

; Sequence 1, Application US/09153599A

```

: Patent No. 6420177
:
: GENERAL INFORMATION:
:
: APPLICANT: Weber, J. Mark
: APPLICANT: Luu, B. Minh
: TITLE OF INVENTION: Method for Strain Improvement of
: TITLE OF INVENTION: Erythromycin Producing Bacterium
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESS: Rockey, Milnamow & Katz, Ltd.
: STREET: 180 N. Stetson Avenue, 2 Prudential Plaza
: CITY: Chicago
: STATE: Illinois
: COUNTRY: U.S.A.
: ZIP: 60601
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/153,599A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mueller, Lisa V.
: REGISTRATION NUMBER: 38,978
: REFERENCE/DOCKET NUMBER: PER159p0041US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5400
: TELEFAX: 312-616-5460
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2299 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: US-09-153-599A-1

```

	Query Match	6.1%;	Score 62.2;	DB 4;	Length 2299;
	Best Local Similarity	48.9%;	Pred. No. 9.3e-11;		
	Matches 197;	Conservative	0; Mismatches 203;	Indels	Gaps 1;
QY	509	TGCTGAACACAGGTTCCTTGGTAGATGCACCTGGTGTTACCAGTAGGAGGAGGAGGAATGG	568		
Db	2296	TCCTGGAGCAGTTGCCGGGACGTC CGGNACAGTGGTGTCCCGCAGCGCGCGGGCTGG	2237		
QY	569	TTGCTGGAAATPAGCATTTACAATTAAAGGCCCTGAAACCTACTGTCAAGGTATACCTGCTG	628		
Db	2236	TCAGCGGCATCGCGCGGCGGTGAAGGCGGAGCACCGCAGGTGCGGGTGCTCGCCGCTCC	2177		
QY	629	AGCCCTCGAATGCAGATGACTGCTACACGCTTAACCTGAAAGAGAAGAACTGACCCCCAATC	688		
Db	2176	AAGCGGAGACGGCGCGCGCTGCGCGCGGTGCTGGCGGGGAAAAACCGGTTGCTCTGTC	2117		
QY	689	TTCAATCCTCCAGAAAACCATAGCAGATGGTGTCAAATCCAGCATTTG---GGTTGAATAACCT	745		
Db	2116	TCGACACCCAGCGGACGATGGCGCCAGCGCATCGCGTCCC CGCGCCGAGCGAGCTGACCT	2057		
QY	746	GGCCTATTATAGAGACCTTGTGGANTGATGCTTTCACTGTACCGGAAGATGAATCAAGT	805		
Db	2056	TGCCCCACGTTCAGGAGCTGCTCGACGACGTGCTACAGTGGGGGAGGAGGCGCTCTCGC	1997		
QY	806	ATGCAACCCACGCTGGTGTGGGGGAGAAATGAACTGCTCATTTGAGCCGACATGCTTGGCGGTG	865		
Db	1996	GCSCGTGCTGCTGTGCC TTGGAACGGCGGAAGCTGGTGGTTCGAACCCGCGCGCTCGCG	1937		
QY	866	CAC TGGCTCGAGTGCTGTCTCAGCATTTCCAACAGTCTCTCC	908		
Db	1936	CGGTCCGCGGACGTCTGGAGCACCCGAGCAGTTTCGGCTCGCC	1894		

## RESULT 4

```

US-08-628-039-1
: Sequence 1, Application US/08628039
: Patent No. 5942660
: GENERAL INFORMATION:
: APPLICANT: Gruys, Kenneth J.
: APPLICANT: Mitsky, Timothy A.
: APPLICANT: Kishore, Ganesh M.
: APPLICANT: Slater, Steven C.
: APPLICANT: Padgett, Stephen R.
: APPLICANT: Stark, David M.
: APPLICANT: Hinchee, Maud A. W.
: APPLICANT: Clemente, Thomas E.
: APPLICANT: Connor-Ward, Dannelte V.
: APPLICANT: Fedele, Mary J.
: APPLICANT: Fry, Joyce E.
: APPLICANT: Howe, Arlene R.
: APPLICANT: Rozman, Renee J.
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and
: TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval
: TITLE OF INVENTION: in Bacteria and Plants
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB
: STREET: 800 No. 5942660th Lindbergh Boulevard
: CITY: St. Louis.
: STATE: Missouri
: COUNTRY: USA
: ZIP: 63167
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/628,039
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE: 13-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Bond, Gary
: REGISTRATION NUMBER: 29,283
: REFERENCE/DOCKET NUMBER: 38-21(13585)A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (314)694-3412
: TELEFAX: (314)695-5435
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1545 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-628-039-1

Query Match 6.0%; Score 60.6; DB 2; Length 1545;
Best Local Similarity 46.5%; Pred. No. 2.5e-10;
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 211 ACGCCAGAGAGAACGCCCAAGCCGTAGTTACTCACAGCAGCGGAACCATGCCCCAAGCT 270
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 220 ACGGAAGAACAGAAAGCGCAGCGGCTGATCACTGCTTCGCGGGTAAACCACGCCGAGGC 279
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 271 CTCACCTATGCTCTAACTGGAAGAAATTCCTGCTTACATTTGTTGTTCCCAACAGCT 330
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 280 GTCGCGCTTTCTCTTCGCGGGTTAGCGGTGAAGGCCCTTGATCGTTTATGCCAACGCCACC 339
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 331 CCCAATGCAAGAAACTGGCAATCCAAAGCCCTATGGAGCATCGATAGTATCTGTGACCCA 390
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 340 GCCGACATCAAAAGTCACCGGCTTCGCGGGCTTCGCGGGGAAGTGTGCTCCACGGCG 399
   ||| |||| |||| | ||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 391 AGTGACGAGTCCAGAGAAAGAGGTCACCTCAAGAAATATTGCAAGAAACAGAACGATCTTG 450

```

Db 400 AACTTTGATGAAGCGAAACGCAAGCGATCGAAGCTGTACAGCAGCAGGCGTTTCACCTGG 459  
QY 451 GTCCATCCCAACAGCAGGCTCGAGTGATAGCTGGACAGGAACAATTGCGCCCTGGGAAGTG 510  
Db 460 GTGCCCGCGTTCACCATCGATCGATGTTGCGCGCAAGCGACGCTGGCGCTGGAACTG 519  
QY 511 CTGAACGAGTTCCTCTTGTGTAGATGCACCTGGTGTACAGTACAGGAGGAGGAATGGTT 570  
Db 520 CTCCAGCAGAGCCCATCTCGACCGCGTATTGTGCCAGTGGCGCGCGCTGTGGCT 579  
QY 571 GCTGGAATAGCCATTACAATTAAAGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGA 629  
Db 580 GCTTGGTGGCGTGCTGATCAAACTGATGCCCAACTCAAACTGATGCCCGGTAGA 638

## RESULT 5

US-08-628-039-5  
; Sequence 5, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannelle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-5

Query Match 6.0%; Score 60.6; DB 2; Length 1545;  
Best Local Similarity 46.5%; Pred. No. 2.5e-10;  
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 211 AGCCAGAGAAGAGCCCAAGCCGTAGTACTACAGCAGCGAAGAACCATGCGCCCAAGCT 270  
Db 220 ACGGAAGAACAGAAAGCGCACGCGGTGATCACTGCTTCTGCGGGTAACACGCGCAGGGC 279  
QY 271 CTACCTATGCTGCTAACTGGAAGAAATTCCTGCTTACATGTTGTTCCCAAAACAGCT 330  
Db 280 GTGCGCTTTTCTGTCGCGGTTAGCGGTGAAGCCCTGATGTTATGCCAACCGCCACC 339  
QY 331 CCCAACTGCAAGAAACTGGCAATCCAAAGCCTATGAGAGCATCGATATATATCTGTGACCCA 390  
Db 340 GCGGACATCAAGTCGACCGGTGCGCGGTTCGCGGGGGAAGTGTGCTTCCACGCGCG 399  
QY 391 AGTGACGAGTCCAGAGAAAGGTCACTCAAGAAATTTATCAAGAAACAGAACGATCTTTG 450  
Db 400 AACTTTGATGAAGCGAAACGCAAGCGATCGAAGCTGTACAGCAGCAGGCGTTTCACCTGG 459  
QY 451 GTCCATCCCAACAGCAGGCGCTCGAGTGATAGCTGGACAGGAACAATTGCGCCCTGGGAAGTG 510  
Db 460 GTGCCCGCGTTCGACCATCCGATGTTGCGCGCAAGCGACGCTGGCGCTGGAACTG 519  
QY 511 CTGAACGAGTTCCTCTTGTGTAGATGCACCTGGTGTACAGTACAGGAGGAGGAATGGTT 570  
Db 520 CTCCAGCAGAGCCCATCTCGACCGCGTATTGTGCCAGTGGCGCGCGCTGTGGCT 579  
QY 571 GCTGGAATAGCCATTACAATTAAAGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGA 629  
Db 580 GCTTGGTGGCGTGCTGATCAAACTGATGCCCAACTGATGCCCGGTAGA 638

## RESULT 6

US-08-628-039-7  
; Sequence 7, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Dannelle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-7

```

; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB
; STREET: 800 No. 5942660th Lindbergh Boulevard
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,039
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE: 13-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bond, Gary
; REGISTRATION NUMBER: 29,283
; REFERENCE/DOCKET NUMBER: 38-21(13585)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)694-3412
; TELEFAX: (314)695-5435
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-628-039-8
;
; Query Match 6.0%; Score 60.6; DB 2; Length 1545;
; Best Local Similarity 46.5%; Pred. No.2.5e-10;
; Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0
;
; QY 211 ACGCCAGACAGAAAGCCCAAGCCGTAGTTACTCACAGACGCGAAACCATGGCCAAAGCT 270
; DB 220 ATCGAGACAGAAAGCGCAGCGGTGATCACTCTTCTCGGGTAACACGCGCAGGGC 279
; QY 271 CTCACCTATGCTGCTAAACTGGAAGAAATTCCTGCTTACATTTGGTTTCCCAACACAGCT 330
; DB 280 GTCGCGTTTCTTCTGCGCGTTAGGCGTGAAGCCCTGATCGTTATGCCAACGCCACC 339
; QY 331 CCCAACTGCAAGAACTGGCAATCCAGCCCTATGGAGCATCGATAGTATCTGTGACCCA 390
; DB 340 GCGGACATCAAGTCGACCGGCTTCGCGGCTTCGCGCGGAAGTGTCTCCACGCGCG 399
; QY 391 AGTGACGAGTCCAGAGAAAGGTGCATCAAGAATTTATGCAAGAAACAGAAAGCATCTTG 450
; DB 400 AACTTTGATGAAGCAAGCAAGCAAGCGATCGAATGTTCACAGACGAGGGGTTACCTTGG 459
; QY 451 GTCATATCCCAACAGGAGCGCTGCACTGATAGCTGGACAAGGAACAATTCGCTTGGAAAGTG 510
; DB 460 GTCCGCGCGTTGACCATCCGATGGTCAATGCCGCGCAAGGCAAGCTGTGCGCTGGAAGTG 519
; QY 511 CTGAACCAAGTTTCCCTTGGTAGATGCACTGGTGTACAGTAGGAGGAGGAGGAATGTTT 570
; DB 520 CTCACAGCAGACGCCCATCTCGACCGGCTAATTTGTGCGAGTCGCGCGCGGCTGTGCT 579
; QY 571 GCTGGAATACCCATTACAATTAAGGCCCTCGAAACCTAGTGTGAAGGTATACGCTCTCA 629
; DB 580 GCTTGGCTGGCGGTGCTGATCAACAACAACTGATGCCGCAAAATCAAGTGATCGCCGTAGA 638
;
; RESULT 8
; US-08-673-388-1
; Sequence 1, Application US/08673388
; Patent No. 5958745
; GENERAL INFORMATION:
; APPLICANT: Gruvs, Kenneth J.

```









```

QY 571 GCTGGGAANTGCCAATTAACAATTAAGGCCCTGAAACCTAGTGTGGAAGGTATACGCTGCTGA 629
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 GCTTGGCTGGCGGTGCTGATCAACAACAACTGATCGCGCAATCAAGGTGATCGCGCTAGA 638

RESULT 14
US-08-614-877-7
; Sequence 7, Application US/08614877
; Patent No. 5959179
; GENERAL INFORMATION:
; APPLICANT: Gruys, Kenneth J.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stark, David M.
; APPLICANT: Hinchee, Maud A. W.
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Connor-Ward, Dannelte V.
; APPLICANT: Fedele, Mary J.
; APPLICANT: Fry, Joyce E.
; APPLICANT: Howe, Arlene R.
; APPLICANT: Rozman, Renee J.
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and
; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval
; TITLE OF INVENTION: in Bacteria and plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F
; STREET: 700 Chesterfield Parkway No. 5959179th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,877
; FILING DATE: 13-MAR-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Charles E.
; REGISTRATION NUMBER: 34,565
; REFERENCE/DOCKET NUMBER: 38-21(10695)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-6224
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-614-877-7

Query Match 6.0%; Score 60.6; DB 2; Length 1545;
Best Local Similarity 46.5%; Pred. No. 2.5e-10;
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 211 ACCCCAGACAGAGAGCCCAAGCCGTAGTTACTCAGCAGCGGAACCATGCGCCAAGCT 270
    ||| |||| |||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 220 ACCGAAGACAGAAAGCGCAGCGCGTGATCTGCTTCGCGGGTAACACGCGCAGGC 279

QY 271 CTCACCTATGCTGCTAAACTGGAAGAAATTCCTGCTTACATTTGTGTCCTCCCAACAGCT 330
    ||| |||| |||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 GTCCGGTTTCTTCTCGCGGGTTAGGCGTGAAGGCCCTGATCTGTTATGCCAACCGCAC 339

QY 331 CCCAACTGCAAGAAACTGGCAATCCAAAGCCTATGGAGCATCGATAGTATACTGTGACCCA 390
    ||| |||| |||| | | | | | | | | | | | | | | | | | | | | | | | | |

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 24, 2003, 04:18:31 ; Search time 21.4683 Seconds  
(without alignments)  
464.608 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISFADVEKAHINQ.....TSLNWVGQARPAPIQTQVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1582.5	90.9	340	4	US-09-789-300A-2
2	365	21.0	424	4	US-09-134-001C-3876
3	359.5	20.7	441	1	US-08-403-866-10
4	303	17.4	436	3	US-08-669-378-2
5	303	17.4	436	3	US-08-669-378-12
6	302	17.4	436	3	US-08-669-378-4
7	302	17.4	436	3	US-08-669-378-6
8	302	17.4	436	3	US-08-669-378-10
9	300	17.2	436	3	US-08-669-378-8
10	296.5	17.0	378	4	US-09-789-300A-4
11	254.5	14.6	325	4	US-09-088-435-1
12	229	13.2	367	4	US-09-134-001C-4168
13	185.5	10.7	551	1	US-08-120-960-2
14	185.5	10.7	551	4	US-09-347-878-9
15	180	10.3	312	4	US-09-134-001C-3920
16	147.5	8.5	319	4	US-09-134-001C-3330
17	128.5	7.4	497	4	US-09-594-193-2
18	125	7.2	487	4	US-09-594-193-13
19	112	6.4	499	4	US-09-594-193-4
20	111.5	6.4	404	4	US-08-887-534A-34
21	104	6.0	434	4	US-09-594-193-7
22	101.5	5.8	484	4	US-09-066-046-8
23	93.5	5.4	403	4	US-09-134-001C-5236
24	93.5	5.4	540	4	US-08-973-462-22
25	93.5	5.4	1786	4	US-08-973-462-8
26	93	5.3	1114	2	US-08-576-626A-31
27	91.5	5.3	619	4	US-09-066-046-2

28	90.5	5.2	212	4	US-08-973-462-25	Sequence 25, Appl
29	90.5	5.2	609	4	US-08-980-115-11	Sequence 11, Appl
30	90.5	5.2	630	4	US-08-973-462-9	Sequence 9, Appl
31	88	5.1	947	4	US-09-228-986-73	Sequence 73, Appl
32	88	5.1	1416	1	US-08-061-465-4	Sequence 4, Appl
33	88	5.1	1642	2	US-08-662-227-2	Sequence 2, Appl
34	88	5.1	1642	4	US-09-017-947-2	Sequence 2, Appl
35	87	5.0	421	2	US-08-576-626A-53	Sequence 53, Appl
36	86	4.9	553	2	US-08-663-566A-13	Sequence 13, Appl
37	86	4.9	553	2	US-08-484-575A-14	Sequence 14, Appl
38	86	4.9	553	2	US-08-023-610-13	Sequence 13, Appl
39	86	4.9	553	2	US-08-288-065A-13	Sequence 13, Appl
40	86	4.9	553	2	US-08-362-240A-13	Sequence 13, Appl
41	86	4.9	553	3	US-08-477-459-14	Sequence 14, Appl
42	86	4.9	553	3	US-08-479-869-14	Sequence 14, Appl
43	86	4.9	553	4	US-08-486-414-14	Sequence 14, Appl
44	86	4.9	553	4	US-08-804-372A-11	Sequence 11, Appl
45	86	4.9	553	5	PCT-US94-01826A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-789-300A-2  
; Sequence 2, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; FILE OF INVENTION: Dependent-Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-789-300A-2

Query Match	90.9%	Score 1582.5	DB 4	Length 340
Best Local Similarity	89.7%	Pred. No. 9.7e-168		
Matches	305	Conservative 19	Mismatches 15	Indels 1
Gaps	1			
Qy	1	MCAQYCISFADVEKAHINQDSIHLPVLTSSILNQIAGRNLPFKCELFQKTSFKIRGA	60	
Db	1	MCAQYCISFADVEKAHINRDSIHLPVLTSSILNQLTGRNLPFKCELFQKTSFKIRGA	60	
Qy	61	LNATIRGLIPDPEKPKAVVTHSSGNHQAITYAAKLEGIPAYIVVPPOTAPNCKKLAIOA	120	
Db	61	LNATIRGLIPDPEKPKAVVTHSSGNHQAITYAAKLEGIPAYIVVPPOTAPNCKKLAIOA	120	
Qy	121	YGASIVYCDPDSERKVTQIMQTEGILVHPNQEPVIAIOGTIALEVLNQPLVDAL	180	
Db	121	YGASIVYCDPDSERKVTQIMQTEGILVHPNQEPVIAIOGTIALEVLNQPLVDAL	180	
Qy	181	VVPVGGGGMVAGIATIKALKPSVKVYAAEPNADDCYQSKLGBELTNLPHPETIADGV	240	
Db	181	VVPVGGGGMVAGIATIKALKPSVKVYAAEPNADDCYQSKLGBELTNLPHPETIADGV	240	
Qy	241	KSSICLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFQT	300	
Db	241	KSSICLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFQT	300	
Qy	301	VSPEVKNCIVLSGNGVDLT-SLNWVGQARPAPIQTQVSV	339	
Db	301	VSPEVKNCIVLSGNGVDLT-SLNWVGQARPAPIQTQVSV	340	

```

; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30, 727
; REFERENCE/DOCKET NUMBER: 20747/30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1600
; TELEFAX: (716) 263-1487
; TELEX: 978450 (WUT)
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Lactococcus lactis subsp. lactis
; INDIVIDUAL ISOLATE: ILVA
;
US-08-403-866-10

Query Match          20.7%; Score 359.5; DB 1; Length 441;
Best Local Similarity 29.4%; Pred. No. 2.9e-31;
Matches 98; Conservative 67; Mismatches 145; Indels 23; Gaps 9;

QY      14 KAHINIQDSIHLPVLTSSILNIGRNIFFKCELFQKTSFKIRGALNAIRGLIPTPE 73
       :|:|:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      34 QANIYLKEVVTKPTQLDPLYSKNKYQANIYLYKEENLOKVRSFKLRGYISIKL---SDE 90

QY      74 EKPXAVTTHSGNHGQALTYAAKLEGIPIAYIVVPOTAPNCKKLIAIQAYGASTIYVC---D 129
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      91 ORSKGVVCASAGNHAOGVAFANQLNISATIEPMTTPNQKISQVKFFGGESHVTIRLGD 150
       ||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY     130 PDESREKVTORTMQETEGILVHPNCEPAVIAGGTTIALEVLAQNP-----LVDAVVYPVG 185
       ||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     151 TDES--ARAKAFFSQNDPKPFIDPFDDENVIACQGVVALEIFQAQAKKGISLDKIFVOIG 209

QY     186 GGMVAGIAITTKALKPSVKVYAEPESNADDCYQSKLKGEITPNLHPPETIADGVK--SSI 244
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     210 GGGLIAGITAYSKERYPQTIELIGVEAKGATSMKAAYSAGO-PVTLKHDKFADGIATV 268

QY     245 GLNTWPIIRDVDVFTVTEDELKAYTQLTVWRMKLLIETAGVALAAYLSQHFOIVSP 304
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     269 GKTTYQLINDKKVQLLAVDEGLISQTILEYSLKLGIVAEPAGATSVAAL-----ELIKDE 323

QY     305 V--KNVCIVLSGNVDLTSLNMVWGQAEPAFY 335
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     324 IKGNKVICIISGNNDISRMOEI--EERALVYE 354

RESULT 4
US-08-669-378-2
; Sequence 2, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-2

Query Match 17.4%; Score 303; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 5.7e-25;
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHHTPVLTSSILNQIAGRNLFKCELFQKTSKIRGALNAIRG 66
Db 21 IRAADIQTAAQRISSVIAPTPLQYCPRLSEETGAETIYLRKREDLQDVRSYKIRGALNS --- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHQALTYAAKLEGIPAYIVVPTAPNCKKLAIOAYG --- 122
Db 78 -GAQSPQEQORDAGIVAASAGNHAQVAYVCKSLGVQGRIVVPVQPKQRDRIMWIGGEF 136

QY 123 ASIVYCDPDSEREKVTQRIMQETEGILVHPNOEPVAVIAGQGTIALEVLNQVPLV ---D 178
Db 137 VSLVVTGNFNDEASAAAHEDAERTGATLIEPEDARNTVIGQGVAAEILSQTSMGKSAD 196

QY 179 ALVVPVGGGGMVAGIATIKALKPSVKVYAAPSNAADDYQSKLKGELTPNLHPETIAD 238
Db 197 HWVPVGGGLLAGVVSVMADMAPRTAIVGIEPAGAAS-MQALHNGGPITLTVDPFVD 255

QY 239 GVK-SSIGLNTWPIIRDLDVDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAAVL 295
Db 256 GAEVKRVGDLNVTIVEKNOGRVHMSATGAVCTEMLDLYQNEGIITAEPAAGALSIAGLKE 315

QY 296 QHFQTVSPVKVNCIVLSGNGVDLTSLNHWGOAER 330
Db 316 MSF---APGSVVVCII-SGGNNDV--LRYAEIAER 344

RESULT 5
US-08-669-378-12
; Sequence 12, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-12

Query Match 17.4%; Score 303; DB 3; Length 436;
Best Local Similarity 28.4%; Pred. No. 5.7e-25;
Matches 95; Conservative 59; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHHTPVLTSSILNQIAGRNLFKCELFQKTSKIRGALNAIRG 66
Db 21 IRAADIQTAAQRISSVIAPTPLQYCPRLSEETGAETIYLRKREDLQDVRSYKIRGALNS --- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHQALTYAAKLEGIPAYIVVPTAPNCKKLAIOAYG --- 122
Db 78 -GAQSPQEQORDAGIVAASAGNHAQVAYVCKSLGVQGRIVVPVQPKQRDRIMWIGGEF 136

QY 123 ASIVYCDPDSEREKVTQRIMQETEGILVHPNOEPVAVIAGQGTIALEVLNQVPLV ---D 178
Db 137 VSLVVTGNFNDEASAAAHEDAERTGATLIEPEDARNTVIGQGVAAEILSQTSMGKSAD 196

QY 179 ALVVPVGGGGMVAGIATIKALKPSVKVYAAPSNAADDYQSKLKGELTPNLHPETIAD 238
Db 197 HWVPVGGGLLAGVVSVMADMAPRTAIVGIEPAGAAS-MQALHNGGPITLTVDPFVD 255

QY 239 GVK-SSIGLNTWPIIRDLDVDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAAVL 295
Db 256 GAEVKRVGDLNVTIVEKNOGRVHMSATGAVCTEMLDLYQNEGIITAEPAAGALSIAGLKE 315

QY 296 QHFQTVSPVKVNCIVLSGNGVDLTSLNHWGOAER 330
Db 316 MSF---APGSVVVCII-SGGNNDV--LRYAEIAER 344

RESULT 6
US-08-669-378-4
; Sequence 4, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:

```



```
US-08-669-378-10
; Sequence 10, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moessel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-10

Query Match 17.4%; Score 302; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 7.3e-25;
Matches 94; Conservative 60; Mismatches 159; Indels 22; Gaps 10;

QY 7 ISFADVEKAHINIQDSIHLPVTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLOYCPRLSEETGAETIYLKREDLQDVRSYKIRGALNS--- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNAQGVAYVCKSLGVQGRYIVVPTPKQRDRIMVHGGEF 136

QY 123 ASIVYCDPDESREKVTQRMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLV----D 178
Db 137 VSLVVTGNFNDEASAAHEDAERTGATLIEPFDARNTVIGQGTVAEIIQSLSMCKSAD 196

QY 179 ALVVPVGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLPHPPTIAD 238
Db 197 HVMPVPGGGLLAGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 255

QY 239 GVK-SSIGLNTWPIIRDLDVDDFTVTEDEIKYATQL--VWGRMKLLIEPTAGVALAVALS 295
Db 256 GAEVKRVGDLNITVEKNOGRVHMSGATEGVCTEMLDLDLYNGBIIAEAPAGALSIALGKE 315

US-08-669-378-8
; Sequence 8, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moessel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-8

Query Match 17.2%; Score 300; DB 3; Length 436;
Best Local Similarity 28.9%; Pred. No. 1.2e-24;
Matches 97; Conservative 58; Mismatches 157; Indels 24; Gaps 11;

QY 7 ISFADVEKAHINIQDSIHLPVTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66
Db 21 IRAADIQTAAQARISSVIAPTPLOYCPRLSEETGAETIYLKREDLQDVRSYKIRGALNS--- 77

QY 67 LIPDTPPEPKA-VVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYG--- 122
Db 78 -GAQSPQEQORDAGIVAASAGNAQGVAYVCKSLGVQGRYIVVPTPKQRDRIMVHGGEF 136

QY 123 ASIVYCDPDESREKVTQRMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLV----D 178
Db 137 VSLVVTGNFNDEASAAHEDAERTGATLIEPFDARNTVIGQGTVAEIIQSLSMCKSAD 196

QY 179 ALVVPVGGGWMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLPHPPTIAD 238
Db 197 HVMPVPGGGLLAGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 255

QY 239 GVK-SSIGLNTWPIIRDLDVDDFTVTEDEIKYATQL--VWGRMKLLIEPTAGVALAVALS 295
Db 256 GAEVKRVGDLNITVEKNOGRVHMSGATEGVCTEMLDLDLYNGBIIAEAPAGALSIALGKE 315
```

Db 197 HVMVPVGGGLAGVWSYMDAPRTAIVGIEPAGAS-MQAALHNGGPITLETVPDFVD 255  
QY 239 G--VKSSIGLNTWPIIRDLYDDVFTVTEDEIKYATOL--VWGRMKLLIEPTAGVALAVAL 294  
Db 256 GGEVKRVGDLN--YIVKNGRVMMSATEGAVCTEMLDIYQNEIIEAPAGALSTAGLK 314  
QY 295 SQHFTQVSPVKNCVIVLSGNGVDLTLNHWGQAE 330  
Db 315 EMSF---APGSVVVCI--SGGNDV--LRYAEIAER 344

RESULT 10  
US-09-789-300A-4  
Sequence 4, Application US/09789300A  
Patent No. 6458576  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
APPLICANT: Rudolph-Owen, Laura A.  
TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
FILE REFERENCE: 35800/208926  
CURRENT APPLICATION NUMBER: US/09/789,300A  
CURRENT FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: US 60/183,208  
PRIOR FILING DATE: 2000-02-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Pyridoxal-Phosphate Dependent Enzyme Family Domain  
OTHER INFORMATION: Sequence  
US-09-789-300A-4

Query Match 17.0%; Score 296.5; DB 4; Length 378;  
Best Local Similarity 30.2%; Pred. No. 2.4e-24;  
Matches 115; Conservative 51; Mismatches 128; Indels 87; Gaps 17;  
QY 19 IQDSIHLPVLTSSILNQIAGRL-----FFKCE-LFQKGTGSPKIRG-ALNAI----- 64  
Db 1 VTELGNTPLVRLNLSKELGEGLGANAAVEIYKLEDLNGPTGSPKDRGLALNMILLAE 60  
QY 65 -----RGLIPDTP-EKPRVAVVTHSSGNHQALTYAAKLEGIPAYIVVQTPAENCKKLA- 117  
Db 61 KLGKGGVGPVQVSKTTIIEPTSGTGIALLAALGLKCTIVMPATDTSREKRAQ 120  
QY 118 IQAYGASIVCDPSDESREKVTQRIQETEGILVHPNQEPV-----IA 161  
Db 121 LRALGAELVVPVAGGSDDLADAIAKAE--LAEENPENAYLLNQAAGFDNPANPEIA 178  
QY 162 GQGTIALEVLNQV-----PLVDALVVPVGGGVMGAGIATIKALKPS-----V 204  
Db 179 GQRTIGPEIWEQGGKEISLGRLPDAVAPVGGGGITGTIARYLKLNPDKGIDVLELPV 238  
QY 205 KYAAEPSNADDCYQSKLKGELT---PN-LH-----PPEIADGVKSSIGLN-- 247  
Db 239 KVIGVPEGS-AVLSGSLKATLTLAGKPGPLHGRDSKYLLODEPVTLPETKSGIGLGVP 297  
QY 248 -----TWPIIROL-----VDDVFTVTEDEIKYATOLVWGRMKLLIEPTAGVALAVALSQ 296  
Db 298 RVGEFVPPILDELLDRROGIDEVVTVTDEALEAARLLAREEGILVGPSSGAVAALKL 357  
QY 297 HFQTVSP--EVKNVCIVLSG 315  
Db 358 AKEGKKPLNKGKTVIVILSG 378

RESULT 11  
US-09-088-435-1  
Sequence 1, Application US/09088435  
Patent No. 6277619

GENERAL INFORMATION:  
APPLICANT: LAL, PREETI  
APPLICANT: CORLEY, NEIL C.  
APPLICANT: GUEGLER, KARL J.  
APPLICANT: PATTERSON, CHANDRA  
TITLE OF INVENTION: SERINE DEHYDRATASE HOMOLOG  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/088,435  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrone, Michael C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0512 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TPL1AZS08  
CLONE: 2752518  
US-09-088-435-1

Query Match 14.6%; Score 254.5; DB 4; Length 325;  
Best Local Similarity 28.1%; Pred. No. 8.8e-20;  
Matches 92; Conservative 59; Mismatches 117; Indels 59; Gaps 15;  
QY 20 QDSIH-LTPVLTSSILNQIAGRLPFKCELFQKGTGSPKIRGALNRLGLIPDTPPEKPKA 78  
Db 11 QEPFHVVTPLLESWALSQVAGMPVFLKCNVQPSGSKIRG-----TGHFCQEMAKKGRH 66  
QY 79 VYTHSSGNHQALTYAAKLEGIPAYIVVQTPAENCKKLAIQAYGASIVYC-----DPSDES 134  
Db 67 LVCSSGCGNAGIAAAYARKLGIPATVLPSTSLQVVQRLQGBGAEVQLTGKVMDEANLR 126  
QY 135 REKVTQRIQETEGILVHPNQEPVAGGTGTALE---VLNQVPLVDALVVPVGGGVMVA 191  
Db 127 AQELAKRQDGNW---VPPFDHPLHWKSHASLVQELKAVLRTPP--GALVLAIVGGGGLA 180  
QY 192 GTAITIKAL-----KPSVKVYAAEPSNADDCYQSKL-KGELTPNLHPPETIAD--GVKSSI 244  
Db 181 GVVAGLLEVGWQHVPIAMHGAH-----CFNNAITAGKLV-----TLPDITSVAKSL 227  
QY 245 GLNT-----WPIIRDLYDDVFTVTEDEIKYATOLVWGRMKLLIEPTAG-VALAA 292  
Db 228 GAKTVAARALECMQVCKIHSEVVEDTEAVS-----AVQOLLDDERMLVEPACCAAAIYS 281  
QY 293 VLSQHPQT---VSPEVKNCIVLSG 316  
Db 282 GLLRRLQAEGLPPSLTSVVIVCGGN 308



RESULT 12  
US-09-134-001C-4168.  
; Sequence 4168, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4168  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4168

Query Match 13.2%; Score 229; DB 4; Length 367;  
Best Local Similarity 26.0%; Pred. No. 7.5e-17;  
Matches 86; Conservative 66; Mismatches 147; Indels 32; Gaps 12;  
QY 26 TPVLTSSILNQIAGRNLFFKCELFQKTSFKIRGALNAIRGLIPDTPEEKPAVTHSSG 85  
DB 44 TPLIYCNLSQKLDIELYKYGANPTGSEKDRGMWA-----VTKAKEQKGVVICASTG 99  
QY 86 N-HQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQ--YGASTVYCDPSDESREKVTQRI 142  
DB 100 NTSASAAVAAR-AGLKAIVPEGKIALGKLS-QAVMYGAETVSTEGNFDEALEIVKEI 157  
QY 143 MQETEGILVHPNQPAVIACQGTIALEVLNQV--PLVDALVVPVGGGMVAG-----IAI 195  
DB 158 AEENDEIELVNSVNPRIEKGKTAFAEIVQDQAPDILAIPVGNAGNITAYWQGFVEY 217  
QY 196 TIKALPSPKVYAAEPSNADDCYQSKLKGELTNLHPPTIADGVKSSIG-LNTWPIIRD 254  
DB 218 HNKNTQLQPMQFGQFQAGSPVQNKI-----INKPETIATAIR--IGNPASWQKAVN 268  
QY 255 LVDD----VFTVTEDEIKYATQLVWGRMKLLIETAGVALAAVLSQHFQTVSPVKNVCI 310  
DB 269 ALDESNGLDSVTEDEILEAYQLMTNEGVSFEPASNASIAGLIKLRSGKLPKGGKIYA 328  
QY 311 VLSGGNV--DLTSLNVWQNERPAPYQTVSV 339  
DB 329 ILTGNGLKDPDTAISLLDNPLOPLPNKESI 359

RESULT 13  
US-08-120-960-2  
; Sequence 2, Application US/08120960  
; Patent No. 5523225  
; GENERAL INFORMATION:  
; APPLICANT: KRAUS, JAN P  
; TITLE OF INVENTION: DNA SEQUENCE ENCODING HUMAN  
; TITLE OF INVENTION: CYSTATHIONINE B-SYNTHASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DILWORTH & BARRESE  
; STREET: 4350 LA JOLLA VILLAGE DRIVE, SUITE 300  
; CITY: SAN DIEGO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/120,960  
; FILING DATE: 12-SEP-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PEPPER PH.D., FREDERICK W.  
; REGISTRATION NUMBER: 31,286  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4410  
; TELEFAX: 619-453-2839  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-120-960-2

Query Match 10.7%; Score 185.5; DB 1; Length 551;  
Best Local Similarity 27.4%; Pred. No. 1e-11;  
Matches 81; Conservative 43; Mismatches 133; Indels 39; Gaps 10;  
QY 26 TPVLTSSILNQIAGR-----NLFFKCELFQKTSFKIRGALNAIRGLIPDTPEEKPAV 80  
DB 87 TPVVR---INKIGKFKGLKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGTILKPGDTII 143  
QY 81 THSSGNHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIVVCD-----PSDES 135  
DB 144 EPTSGNTGIGLALAAAVRGYRCIIIVPEKMSSEKVDVLRALGAIEIVRTPTNARFDSPE 203  
QY 136 EKVQRMOP--TEGILVHPNQPAVIACQGTIALEVLNQV--LVDAVVPVGGGMVAG 192  
DB 204 VGVWRLKNEIPNSHILDOVRNASNPLAHYDTTDAETLQCCDKGLMDLVAASVGTGGTITG 263  
QY 193 IAITIKALPSPKVYAAEPSNADDCYQSKLKGELTNLHPPTIADGVKSS-----IGLN 247  
DB 264 IARKLEKPCGRIIGVDP-----EGSI---LAEPELNQTEOTTYYEVEGYD 309  
QY 248 TWPIL--RLVDVDFVTEDEIKYATQLVWGRMKLLIETAG--VALAAVLSQHFQ 299  
DB 310 FIPVLDRTVVDVWFKNSDEEAFTRMIAQOGLCGGSAGSTVAVVAAQELQ 365

RESULT 14  
US-09-347-878-9  
; Sequence 9, Application US/09347878C  
; Patent No. 6376210  
; GENERAL INFORMATION:  
; APPLICANT: Yuan, Chong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES  
; FILE REFERENCE: 25885-1651  
; CURRENT APPLICATION NUMBER: US/09/347,878C  
; CURRENT FILING DATE: 1999-07-06  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-347-878-9

Query Match 10.7%; Score 185.5; DB 4; Length 551;  
Best Local Similarity 27.4%; Pred. No. 1e-11;  
Matches 81; Conservative 43; Mismatches 133; Indels 39; Gaps 10;  
QY 26 TPVLTSSILNQIAGR-----NLFFKCELFQKTSFKIRGALNAIRGLIPDTPEEKPAV 80  
DB 87 TPVVR---INKIGKFKGLKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGTILKPGDTII 143  
QY 81 THSSGNHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIVVCD-----PSDES 135  
DB 144 EPTSGNTGIGLALAAAVRGYRCIIIVPEKMSSEKVDVLRALGAIEIVRTPTNARFDSPE 203

```
OY 136 EKVTRIMQE--TEGILVHPNQPAVIAGOGTTALEVLNVP-LVDALVVPVGGGGMVAG 192
Db 204 VGVAWRLKNEIPNSHLDQYRNASNPLAHYDTTADETLQCCDKGLDMLVASVGTGGTTG 263
OY 193 IAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTNLHPETIADGVKSS-----IGLN 247
Db 264 IARKLEKCPGCRIGVDP-----EGSI---LAPEELNOTEQTYTEVEGIGYD 309
OY 248 TWPII--RLVDVDFVTTEDEIKYATQLVWGRMKLLIEPTAG--VALAAVLSOHFO 299
Db 310 FIFTVLDRTVVDKWFKNDEEAEFTFARMLIAEQELLCGSGSNGSTVAVAKAAELQ 365

RESULT 15
US-09-134-001C-3920
: Sequence 3920, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3920
: LENGTH: 312
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3920
```

```
Query Match: 10.3%; Score 180; DB 4; Length 312;
Best Local Similarity 25.0%; Pred. No. 1.6e-11;
Matches 75; Conservative 49; Mismatches 140; Indels 36; Gaps 10;

OY 21 DSHLTPVLTSSILNOIAGRN--LFPKCELFQKTSFKIRGALNAIRGLIPDTPPEKPKA 78
Db 15 DLIGQPLV---LLESFSDENVKIYAKLEQFNPGSGSIKDRGLYKLEKAIDEGRKREGDT 71

OY 79 VVTHSSGNHQALTYAAKLEGIPAYIVPQTPAPNCKKLAIQAYGASIVYCDPSDESREKV 138
Db 72 IVEATAGTGTGGLAIASNRHKVKCIIFAPGFAEEKISIMKALGADVRRTPRAEGMTGAQ 131

OY 139 TORIMQETEGILVHPNQ-----EPAVIAGOGTTALEVLNVPVLDALVVPVGGGGMVAGI 193
Db 132 OEALAYATRYGYLYMNQFETKDNPG--AVTQTLAKQITDELSHIDYFVAGVSGGTTFTGV 189

OY 194 AITIKALKPSVKVYAAEPSNADDCYQSKLKGELTNLHPETIADGVKSSIGLNTWP--I 251
Db 190 AQHLKTY--DVKNYIYEPEG-----SVLNGGVV---HPHAT-----EGIGSEKWPFSFL 232

OY 252 IRDLVDVDFVTTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSOHFOQVSPVKNVCIV 311
Db 233 EKELVDGIFTVDKDAFNKVNKLVANKEGLLVGSSGSAALQALE-----LKKSIQNGVIV 287
```

Search completed: June 24, 2003, 05:36:43  
Job time : 44.4683 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 07:02:30 ; Search time 1462 Seconds  
(without alignments)  
11332.410 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgcattcat.....atcagctgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	803	78.5	1398	11 AK017235	AK017235 Mus muscu
2	751	73.4	785	13 BI763753	BI763753 603047780
3	729.2	71.3	848	9 AU131397	AU131397 AU131397
4	602	58.8	634	14 BM719814	BM719814 UI-E-EJ0-
5	526	51.4	527	9 AA446793	AA446793 zw89f02.r
6	523.8	51.2	769	12 BF163794	BF163794 601769701

7	491.8	48.1	667	10	BB609829	BB609829
8	487.6	47.7	652	10	AV709023	AV709023
9	479.8	46.9	600	13	BI988879	BI988879
10	479.2	46.8	697	10	AV649313	AV649313
11	451	44.1	819	13	BI739030	BI739030
12	450.8	44.1	565	10	BE655084	BE655084
13	410.4	40.1	1159	14	BQ215405	BQ215405
14	402.4	39.3	918	14	BQ885621	BQ885621
15	391.8	38.3	823	13	BI738966	BI738966
16	383	37.4	391	14	BQ303565	BQ303565
17	380.6	37.2	404	14	BQ303568	BQ303568
18	379.6	37.1	386	14	BQ303566	BQ303566
19	376.6	36.8	698	9	AV022510	AV022510
20	365.4	35.7	455	9	AI332578	AI332578
21	363.8	35.6	602	10	AV374491	AV374491
22	352.8	34.5	471	14	H73097	H73097
23	348.6	34.1	729	13	BC965678	BC965678
24	341	33.3	394	10	AV649514	AV649514
25	337	32.9	793	12	BG404240	BG404240
26	329	32.2	742	9	AI525507	AI525507
27	328.2	32.1	827	13	BI332919	BI332919
28	323.8	31.7	438	14	H86748	H86748
29	321.2	31.4	519	10	BE641631	BE641631
30	320.8	31.4	830	12	BF135412	BF135412
31	315.4	30.8	556	10	BB692486	BB692486
32	300.2	29.3	380	9	AA034539	AA034539
33	299	29.2	661	10	BB621410	BB621410
34	295.4	28.9	688	10	BB625125	BB625125
35	289.8	28.3	1087	12	BF532321	BF532321
36	282	27.6	544	9	AA432108	AA432108
37	275.2	26.9	645	10	AW211242	AW211242
38	274.6	26.8	646	10	BB644584	BB644584
39	274	26.8	566	14	BM766019	BM766019
40	273	26.7	288	14	243188	243188
41	266.4	26.0	632	10	BB641468	BB641468
42	266	26.0	611	10	BB618702	BB618702
43	265.4	25.9	657	10	AW319255	AW319255
44	258.2	25.2	672	10	BE380996	BE380996
45	255.6	25.0	426	14	D53895	D53895

ALIGNMENTS

RESULT 1	AK017235	1398 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK017235	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330405D10:serine racemase, full insert sequence.			
ACCESSION	AK017235				
VERSION	AK017235.1	GI:12856372			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:5330405D10.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	Carninci, P. and Hayashizaki, Y.			
AUTHORS	High-efficiency full-length cDNA cloning				
TITLE	Meth. Enzymol. 303, 19-44 (1999)				
JOURNAL	99279253				
MEDLINE	10349636				
PUBMED					
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
TITLE	Genome Res. 10 (10), 1617-1630 (2000)				
JOURNAL	20499374				
MEDLINE	11042159				
PUBMED					



Db 890 AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACAGTCT 949  
 QY 661 AAGCTGAAGGGAAACTGATGCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCTC 720  
 Db 950 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTCTC 1009  
 QY 721 AAATCCAGCATTTGGCTTGAACACCTGGCTATTATCAGGACCTTGTGGATGATATCTTC 780  
 Db 1010 AAATCCAGCATTTGGCTTGAATACCTGGCTATTATAGAGACCTTGTGGATGATGCTCTTC 1069  
 QY 781 ACTGTACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGGATGAAACTA 840  
 Db 1070 ACTGTACAGAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGAGATGAAGCTG 1129  
 QY 841 CTCAFTGAACCTACAGCTGGTGTGGAGTGGCTGCTGTCTCAACATTTTCAAACT 900  
 Db 1130 CTCAFTGAGCGACTGCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189  
 QY 901 GTTTCCTCCAGAGTAAAGAACTTTGTATTGTCTCAGTGTGGTGAATGTAGACTTAACC 960  
 Db 1190 GTCTCTCCAGAGTAAAGAACTTTGTATTGTCTCAGTGTGGTGAATGTAGACTTAACC 1247  
 QY 961 TCCCTCCATAAATCTGGGTGAAGCAGCTGAAAGGCCAGCTTCTTATFACAGTCTGTTTCTGTT 1020  
 Db 1248 -CCTCCCTGAACCTGGGTGGGCGAGCTGAGCGGAGCTCTTACCAGCGTTTCTGTT 1306  
 QY 1021 TAA 1023  
 Db 1307 TAA 1309

## RESULT 2

BI763753

LOCUS

DEFINITION 603047780F1 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5188099 5',  
 mRNA sequence.

ACCESSION

BI763753

VERSION

BI763753.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 785)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M11469 row: p column: 20

High quality sequence stop: 783.

Location/Qualifiers

1. 785

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5188099"

/clone\_lib="NIH\_MGC\_116"

/lab\_host="DH10B"

/note="Organ: pooled colon, kidney, stomach; Vector:

pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA

source anonymous pool of 3 colons, age 26 yo male, 49 yo

female, 71 yo male colon; 46 yo male kidney, and pool of 2

stomachs, 62 yo male and 70 yo female. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.4 kb,

insert size range 1-3 kb. Library is normalized and

## FEATURES

source

enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 023. Note: this is a NIH\_MGC Library."  
 BASE COUNT 225 a 168 c 200 g 192 t  
 ORIGIN

Query Match 73.4%; Score 751; DB 13; Length 785;

Best Local Similarity 99.6%; Pred. No. 1.7e-214;

Matches 784; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 104 ATCAACTAACAGGGCGCAATCTTTTCTCAATGTGAATCTTCCAGAAAACAGCATCTT 163  
 Db 1 ATCAACTAACAGGGCGCAATCTTTTCTCAATGTGAATCTTCCAGAAAACAGCATCTT 60  
 QY 164 -TTAAGATTCTGGTGTCTCTCAATGCCCTCAGAACTTTGGTTCCTGTATGCTTTAGAAAG 222  
 Db 61 GTTAAGATTCTGGTGTCTCTCAATGCCCTCAGAACTTTGGTTCCTGTATGCTTTAGAAAG 120  
 QY 223 AAGCGGAAGCTGTTGTACTACAGCAGTGGAAACCATGGCCAGGCTCTCACTATGCT 282  
 Db 121 AAGCGGAAGCTG-TGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCACTATGCT 179  
 QY 283 GCCAAATTGGAAGGAATTCCTGCTTATATTGTGGTGGCCAGACAGCTCCAGACTGTAAA 342  
 Db 180 GCCAAATTGGAAGGAATTCCTGCTTATATTGTGGTGGCCAGACAGCTCCAGACTGTAAA 239  
 QY 343 AAATTTGCAATACAAAGCCTACGGAGCTCAATTGTATATCTGTGAACCTAGTATGAGTCC 402  
 Db 240 AAATTTGCAATACAAAGCCTACGGAGCTCAATTGTATATCTGTGAACCTAGTATGAGTCC 299  
 QY 403 AGAGAAATTTGCAAAAGAGTTACAGAAAGAAAGAGGATCATGGTATCATCCCAAC 462  
 Db 300 AGAGAAATTTGCAAAAGAGTTACAGAAAGAAAGAGGATCATGGTATCATCCCAAC 359  
 QY 463 CAGGAGCTGCTGAGTATAGCTGGCAAGGACAATTCCTGGAAAGTCTGTAACCAAGTT 522  
 Db 360 CAGGAGCTGCTGAGTATAGCTGGCAAGGACAATTCCTGGAAAGTCTGTAACCAAGTT 419  
 QY 523 CCTTTGGTGGATGCACTGGTGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCA 582  
 Db 420 CCTTTGGTGGATGCACTGGTGTACTGTAGTGGAGGAGGAATGCTTGTGGAATAGCA 479  
 QY 583 ATTACAGTTAAGGCTCTCAAACTAGTGTGAAGGTATATGCTGTGAACCTCAATGCA 642  
 Db 480 ATTACAGTTAAGGCTCTCAAACTAGTGTGAAGGTATATGCTGTGAACCTCAATGCA 539  
 QY 643 GATGACTCTACCACTGCAAGCTGAAGGGAAGAACTGATGCCCAATCTTTATCTCCAGAA 702  
 Db 540 GATGACTCTACCACTGCAAGCTGAAGGGAAGAACTGATGCCCAATCTTTATCTCCAGAA 599  
 QY 703 ACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTTGAACACCTGGCCCTATTATCAGGAC 762  
 Db 600 ACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTTGAACACCTGGCCCTATTATCAGGAC 659  
 QY 763 CTTGTGGATGATATCTTCACTGTACAGAGATGAATTAAGTGTGCAACCCAGCTGGTG 822  
 Db 660 CTTGTGGATGATATCTTCACTGTACAGAGATGAATTAAGTGTGCAACCCAGCTGGTG 718  
 QY 823 TGGGAGAGGATGAAGAACTACTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTG 882  
 Db 719 TGGGAGAGGATGAAGAACTACTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTG 778  
 QY 883 TCTCAAC 889  
 Db 779 TCTCAAC 785

## RESULT 3

AUI31397

LOCUS

DEFINITION

AUI31397 NT2RP3 Homo sapiens cDNA clone NT2RP3002501 5', mRNA

sequence.

ACCESSION

AUI31397

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
AU131397.1	GI:10991751	EST.	Human.						
			Homo sapiens						
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
			1 (bases 1 to 848)						
			Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.						
			HRI human cDNA project						
			Unpublished (2000)						
			Contact: Takao Isogai						
			Genomics Laboratory						
			Helix Research Institute						
			1532-3 Yana, Kisarazu, Chiba 292-0812, Japan						
			Tel: 81-438-52-3975						
			Fax: 81-438-52-3986						
			Email: genomics@hri.co.jp						
			HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute						
			Location/Qualifiers						
			1. .848						
			/organism="Homo sapiens"						
			/db_xref="taxon:9606"						
			/clone_lib="NT2RP302501"						
			/cell_type="teratocarcinoma"						
			/cell_line="NT2"						
			/note="Vector: pME185FL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"						
			BASE COUNT 236 a 187 c 214 g 207 t 4 others						
			ORIGIN						
			Query Match 71.3%; Score 729.2; DB 9; Length 848;						
			Best Local Similarity 98.0%; Pred. No. 6.4e-208;						
			Matches 768; Conservative 0; Mismatches 12; Indels 4; Gaps 3;						
			QY 1 ATGTGTCACGATGTCATCTCTCTGATGTTGAAAAGCTCATCAACATCGA 60						
			DB 69 ATGTGTCACGATGTCATCTCTCTGATGTTGAAAAGCTCATCAACATCGA 128						
			QY 61 GATTTCTATCCAGCTCACACGAGTGTAAACAGCTCCATTTGAATCAACTAACAGGCGC 120						
			DB 129 GATTTCTATCCAGCTCACACGAGTGTAAACAGCTCCATTTGAATCAACTAACAGGCGC 188						
			QY 121 ATCTTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCGTTGGTCT 180						
			DB 189 AATCTTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATTCGTTGGTCT 248						
			QY 181 CTCATGCGCTCAGAGCTTGTCTCTGATGTTAGAAAGGAGCCGAAAGCTGTGTT 240						
			DB 249 CTCATGCGCTCAGAGCTTGTCTCTGATGTTAGAAAGGAGCCGAAAGCTGTGTT 308						
			QY 241 ACTCACAGAGTGGAAACACCTGCGCAGGCTCTCACCTATGCTGCCAAATTCGAAGGAAT 300						
			DB 309 ACTCACAGAGTGGAAACACCTGCGCAGGCTCTCACCTATGCTGCCAAATTCGAAGGAAT 368						
			QY 301 CTGCTTATATTTGTTGGTCCCGCAGACGCTCCAGACTGTAAACAACTTGCATAACAGCC 360						
			DB 369 CTGCTTATATTTGTTGGTCCCGCAGACGCTCCAGACTGTAAACAACTTGCATAACAGCC 428						
			QY 361 TACGAGCGCTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAATTTGTCRAAA 420						
			DB 429 TACGAGCGCTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAATTTGTCRAAA 488						
			QY 421 AGAGTTTACAGAAGAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 480						
			DB 489 AGAGTTTACAGAAGAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 548						
			QY 481 GCTGCAGAGGGACAATTCGCTGGAGTGTCTGTAACAGGTTCTTTGGTGGATGCACGTG 540						



QY 454 CATCCCAACGAGCGCTGCACTGATAGTGGACAAAGGACAAATTCCTCGCAAGTGTG 513  
 |||||||  
 Db 422 CATCCCAACGAGCGCTGCACTGATAGTGGACAAAGGACAAATTCCTCGCAAGTGTG 481  
 |||||||  
 QY 514 AACCAAGGTTCCCTTTGGTGGATGCACTGGTGGTACCTGTAGTGGAG 559  
 |||||||  
 Db 482 AACCAAGGTTCCCTTTGGTGGATGCACTGGTGGTACCTGTAGTGGAG 527  
 |||||||

RESULT 6  
 BF163794  
 LOCUS 601769701f1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3988873 5',  
 mRNA sequence. EST 30-OCT-2000

ACCESSION BF163794  
 VERSION BF163794.1 GI:11043971

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota

REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 769)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14W9198 row: a column: 02

High quality sequence stop: 699.

Location/Qualifiers

1..769

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:3988873"

/clone\_lib="NCI\_CGAP\_Lu29"

/tissue\_type="spontaneous tumor, metastatic to mammary."

Stem cell origin.

/lab\_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;

Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH

BASE COUNT 225 a 186 c 187 g 171 t

ORIGIN

Query Match 51.2%; Score 523.8; DB 12; Length 769;  
 Best Local Similarity 87.1%; Pred. No. 3.8e-146;  
 Matches 633; Conservative 0; Mismatches 87; Indels 7; Gaps 5;

QY 1 ATGTGCTGCTAGTATGCTATCCCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
 |||||||  
 Db 40 ATGTGCTGCTAGTATGCTATCCCTTCTGCTGATGTTGAAAAGCTCATATCAACATTCAA 99  
 |||||||

QY 61 GATTCATCCACCTCACACCACTGCTTAACAGCTCCATTTTGAATCAACTAACAGGCGC 120  
 |||||||  
 Db 100 GATTCATCCACCTCACACCACTGCTTAACAGCTCCATTTTGAATCAACTAACAGGCGC 159  
 |||||||

QY 121 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAAGATTGCTGGTCT 180  
 |||||||  
 Db 160 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAAGATTGCTGGTCT 219  
 |||||||

QY 181 CTCATGCGCTCAGAGCTTGGTTCCTGATGC-TTTAGAAAGGAGCCCAAGCTGTTGT 239  
 |||||||  
 Db 220 CTTAATGCCATCAGAGCTTAAATCTGACACGCAAGAAAGAGCCCAAGCGGTAGT 279  
 |||||||

QY 240 TACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCCTATGCTGCCAAATTTGGAAGGAAT 299  
 |||||||  
 Db 280 TACTCACAGCAGCGGAACCATGGCCAGGCTCTCACCCTATGCTGCCAAATTTGGAAGGAAT 339  
 |||||||  
 QY 300 TCCTGCTTTATATTGTGGTGGCCCGACAGCTCCAGAGCTGTAATAAATTTGCAATCAAGC 359  
 |||||||  
 Db 340 TCCTGCTTTATATTGTGGTGGTTCCTCAACAGCTCCCAACTGCAAGAAATTTGCAATCAAGC 399  
 |||||||  
 QY 360 CTACGAGCGTCATATTGTATATTGTAACCTAGTGTAGTCCAGAGAAATTTGTCATAA 419  
 |||||||  
 Db 400 CTATGGAGCATCGATAGTATATCTGTGACCCAAAGTGACGAGTCCAGAGAAATTTGTCATAA 459  
 |||||||  
 QY 420 AAGAGCTTACAGAAAGAACAGAGGAGCATGCTGATCATCCCAACAGGAGCCTGCAAGTAT 479  
 |||||||  
 Db 460 AAGAATTATCAAGAAAGAACAGAGGAGCATCTTGGTCTCCCAACAGGAGCCTGCAAGTAT 519  
 |||||||  
 QY 480 AGCTGGACAAGGAGCAATTTGGCTGGAAGTGTGAACCAAGGTTCTTTTGGTGGATGCAT 539  
 |||||||  
 Db 520 AGCTGGACAAGGAGCAATTTGGCTGGAAGTGTGAACCAAGGTTCTTTTGGTGGATGCAT 579  
 |||||||  
 QY 540 GGTGTGCTAGTGTAGTGGAGGAGGAATGCTTGTGCTGGAATAGCAATTTAGAGCTCT 599  
 |||||||  
 Db 580 GGTGTGCTAGTGTAGGAGGAGGAATGCTTGTGCTGGAATAGCAATTTAGAGCTCT 638  
 |||||||  
 QY 600 GAAACCTAGTGTGAAGTATATGCTGCTGAACCTCAATGCAGATGCTGCTACCAAGTC 659  
 |||||||  
 Db 639 GAAACCTAGTGTGAAGTATATGCTGCTGGAACCTCAATGCAGATGCTGCTACCAAGTC 697  
 |||||||  
 QY 660 CAAGCTGAAGGGAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGAGATGGTGT 719  
 |||||||  
 Db 698 TAACTGAAGGAGAGAACTG--GCCCAATTTAATCTCAAGAAAC--ATAGCGATGGTGT 753  
 |||||||

QY 720 CAAATCC 726

Db 754 CAAATCC 760

RESULT 7

BB609829

LOCUS

DEFINITION

BB609829

ACCESSION

BB609829

VERSION

BB609829.1

KEYWORDS

EST

SOURCE

house mouse.

Mus musculus

ORGANISM

REFERENCE

1 (bases 1 to 667)

AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,

, Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakai, H., Sasaki

, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

CONTACT: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

URL: <http://genome.gsc.riken.go.jp/>

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp)

Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh

, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new





```

Qy 181 CTCATGCCCTCAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCGAAGAGCTGTGTT 240
    |||||
Db 236 CTCATGCCCTCAGAGCTTGGTTCCTGATGCTTTAGAAAGAGCGAAGAGCTGTGTT 295

Qy 241 ACTCAGCAGAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATGGGAAGAA 300
    |||||
Db 296 ACTCAGCAGAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATGGGAAGAA 355

Qy 301 CCTGCTTATATCTGTGTGCCCCAGACAGCTCCAGAGCTGTAAGAACTGCAATACAGCC 360
    |||||
Db 356 CTTGCTTATATCTGTGTGCCCCAGACAGCTCCAGAGCTGTAAGAACTGCAATACAGCC 415

Qy 361 TACGGAGCGCTCAATTTGATCTGTAACCTAGTAGTGATGAGTCCAGAGAAATGTTCAAAA 420
    |||||
Db 416 TACGGAGCGCTCAATTTGATCTGTAACCTAGTAGTGATGAGTCCAGAGAAATGTTCAAAA 472

Qy 421 AGATTACAGAGAAACAGAGGCATCATGGTACATCCCAACAGGAGCGCTGCAGTGA 480
    |||||
Db 473 -----GCAGTCATA 481

Qy 481 GCTGGACAGGAGCAATTCCTCGAAGTCTCGAACCAAGCTTCTTTGGTGGATGCAC 540
    |||||
Db 482 GCTGGACAGGATCAATTCCTCGAAGTCTCGAACCAAGTCTTTGGTGGATGCAC 541

Qy 541 GTGGTACCTGTAGGTGGAGGAGGAATGCTTGTGGAATAGCAATTAAGTAAAGGCTCTG 600
    |||||
Db 542 GGGGTACCTGTAAGTGGAGGATCAATGCTTGTGGAATTAAGTAAAGGCTCTG 600

Qy 601 AATCTGTAGTGAAGTATATGCTGCTGACCTCAATGCAATGCAATGCAATGCAATGCA 652
    |||||
Db 601 AATCTGTAGTGAAGTATATGCTGCTGACCTCAATGCAATGCAATGCAATGCAATGCA 651

```

## RESULT 9

```

BI988879
LOCUS 4022-49 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
      mRNA sequence.
ACCESSION BI988879
VERSION BI988879.1 GI:17959868
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
        White, R.A., Beremand, P.D., Thomas, F.L., Gan, L. and Klein, W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing
        and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
COMMENT Contact: Klein WH
        Department of Biochemistry and Molecular Biology
        University of Texas M.D. Anderson Cancer Center
        Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
        Tel: 713 792 3646
        Fax: 713 790 0329.

```

## FEATURES

```

Source Location/Qualifiers
1..600
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
   /tissue_type="neural retina"
   /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
   Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
   (Manniat); Cloning Technique: CUA Cloning (Clontech),
   Life Technologies; Average insert size: 1.8 Kb;
   Insertion site: TACGTCACCTGAATTCGAGTG---. Other
   information regarding entire library may be found at
   http://pga.swmed.edu/data/Libraries/microarray_cdna_librar

```

```

BASE COUNT 177 a 145 c 134 g 136 t 8 others
ORIGIN
Query Match 46.9%; Score 479.8; DB 13; Length 600;
Best Local Similarity 87.4%; Pred. No. 5.9e-133;
Matches 520; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1 ATCTGTGCTCAGTATTCGCATCTCTTTCCTGATGCTTTGAAAAAGCTATATACACATTCGA 60
    |||||
Db 6 ATGTGTGCTCAGTATTCGCATCTCTTTCCTGATGCTTTGAAAAAGCTATATACACATTCGA 65

Qy 61 GATTCTATCCACCTCAGACAGCTGCTAACAGCTCCATTTTGAATCAACCTAACAGGCGC 120
    |||||
Db 66 GACTNTATCCACCTCAGACAGCTGCTAACAGCTCCATTTTGAATCAANAGCAGGCGC 125

Qy 121 AATCTTTTCTCAATTTGTAACCTCTTCCAGAAACAGGATCTTTTAAAGTTCGTGGTGT 180
    |||||
Db 126 AATCTTTTCTCAATTTGTAACCTCTTCCAGAAACAGGATCTTTTAAAGTTCGTGGTGT 185

Qy 181 CTCAATGCCGTGCAAGCTTGGTTCCTGATGCTTTAGAAAGAGCGGAAAGCTGTGTT 240
    |||||
Db 186 CTTAATGCCATCAGAGCTTAACTCTGACACGCCAGAGAGAGCCCAAGCCGTAGTT 245

Qy 241 ACTCAGCAGAGTGAACCACTGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAA 300
    |||||
Db 246 ACTCAGCAGAGTGAACCACTGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGAA 305

Qy 301 CCTGCTTATATTTGTTGGTGGCCAGACAGCTCCAGACTGTAAACAACTTGCATATACAGCC 360
    |||||
Db 306 CCTGCTTATATTTGTTGGTGGCCAGACAGCTCCAGACTGTAAACAACTTGCATATACAGCC 365

Qy 361 TACGGAGCGCTCAATTTCTATCTGTAACCTAGTAGTGATGAGTCCAGAGAAATTTGCAAAA 420
    |||||
Db 366 TATGGAGCATCGGTAGTATCTGTGACCAAGTGCAGAGTCCAGAGAAAGGTCACTCAA 425

Qy 421 AGATTACAGAGAAACAGAGGATCATGGTATACATCCCAACAGGAGCGCTGCAGTGA 480
    |||||
Db 426 AGAATTTATCAAGANACAGAGGATCTTGGTCCATCCCAACAGGAGCGCTGCAGTGA 485

Qy 481 GCTGGACAGGACAAATTTGCCCTGGAGTGTCTGAACAGGATCTCTTTGGTGGATGAC 540
    |||||
Db 486 GCTGGACAGGACAAATTTGCCCTGGAGTGTCTGAACAGGATCTCTTTGGTGGATGAC 545

Qy 541 GTGGTACCTGTAGGTGAGGAGGAATGCTTGTGGAATAGCAATTAAGTAAAG 595
    |||||
Db 546 GTGGTACCTGTAGGTGAGGAGGAATGCTTGTGGAATAGCAATTAAGTAAAG 600.

```

## RESULT 10

```

AV649313
LOCUS AV649313 GLC Homo sapiens cDNA clone GLCBSF06 3', mRNA sequence.
DEFINITION AV649313
ACCESSION AV649313
VERSION AV649313.1 GI:9870327
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 697)
Xiao, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z., and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106

```

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

## CONTACT

201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzgchgc.sh.cn  
This clone is available at CHGC in Shanghai.

## FEATURES

Location/Qualifiers  
1..697

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="GLCSF06"  
/clone\_lib="GLC"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 190 a 154 c 174 g 171 t 8 others

Query Match 46.8%; Score 479.2; DB 10; Length 697;  
Best Local Similarity 87.1%; Pred. No. 9.5e-133;  
Matches 601; Conservative 0; Mismatches 31; Indels 58; Gaps 5;

QY 1 ATGTGTCAGTATTGTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
DB 54 ATGTGTCAGTATTGTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 113  
QY 61 GATTCTATCCACCTCACACAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
DB 114 GATCTATCCACCTCACACAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 173  
QY 121 AATCTTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAAGATCGTGGTCT 180  
DB 174 AATCTTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAAGATCGTGGTCT 233  
QY 181 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGTT 240  
DB 234 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGCCGAAAGCTGTTGTT 293  
QY 241 ACTCACAGCTGGAACCATGGCCAGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 300  
DB 294 ACTCACAGCTGGAACCATGGCCAGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 353  
QY 301 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 360  
DB 354 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 413  
QY 361 TACGAGGCGTCAATGTATATCTGTGAACCTTAGTATGATGATGATGATGATGATGAT 420  
DB 414 TACGAGGCGTCAATGTATATCTGTGAACCTTAGTATGATGATGATGATGATGATGAT 459  
QY 421 AGAGTTACAGAAACAGAGGATCATGTTACATCCCAACAGGAGCCCTGCAGTGATA 480  
DB 460 -----ACCAGGAGCCCTGCAGTGATA 479  
QY 481 GCTGACAAAGGACAATTCCTCGAAGTGTGAACAGGTTCTCTTTGGTGGATGCACTG 540  
DB 480 GCTGACAAAGGACAATTCCTCGAAGTGTGAACAGGTTCTCTTTGGTGGATGCACTT 539  
QY 541 GTGGTACCTGTAGGTGGAGGAGGATGCTTCTGGAATAGCAATTTACAGTTAAGGCTCTG 600  
DB 540 GTGGTACCTGTAGGTGGAGGAGGATGCTTCTGGAATAGCAATTTACAGTTAAGGCTCTG 598  
QY 601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATTCAGATGACTGTACCAAGTCC 660  
DB 599 AAA-CTAGTGTGAAGG-ATATGCTGTGAACCTTC-AATGCANATGACTGCTACCAAGTC 655  
QY 661 AAGCTGAAGGGAACTGATGCCCAATCTT 690  
DB 656 AAGCTGAAGGGAACTGATGCCCAATCTT 685

RESULT 11

BI739030

LOCUS

603359880F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5367042 5',  
mRNA sequence.

ACCESSION

BI739030

VERSION

BI739030.1 GI:15716056

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 819)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11933 row: p column: 19

High quality sequence stop: 817.

FEATURES

source

1..819  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5367042"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT

236 a 208 c 192 g 183 t

ORIGIN

Query Match 44.1%; Score 451; DB 13; Length 819;  
Best Local Similarity 89.0%; Pred. No. 3.1e-124;  
Matches 487; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGTGTCAGTATTGTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
DB 273 ATGTGTCAGTATTGTCATCTCTTCTGCTGATGTTGAAAAGCTCATATCAACATTGCA 332  
QY 61 GATTCTATCCACCTCACACAGTCTAACAGCTCCATTTTCAATCAACTAACAGGGCGC 120  
DB 333 GACTCTATCCACCTCACCCAGTGTGTAACAGCTCCATTTTGAATCAATACAGGGCGC 392  
QY 121 AATCTTTTCTTCAATGTGAATCTTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT 180  
DB 393 AATCTTTTCTTCAATGTGAGCTTCCAGAAACAGTGGTCTTTTAAAGATTCAGGTGCC 452  
QY 181 CTCAATGCGCTCAGAGCTTGGTCTCTGATGCTTTTGAAGAGGAGCCGAAAGCTGTTGT 240  
DB 453 CTTAATGCATCAGAGGCTTAATCTCTGACAGCCAGAGAGAGAGCCCGTAGTT 512  
QY 241 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 300  
DB 513 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 572  
QY 301 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 360  
DB 573 CTGCTTATATGTGGTGGCCCGACAGCTCCAGACTGTAAACAACTTGCATACAGCC 632  
QY 361 TACGAGGCGTCAATGTATATCTGTGAACCTTAGTATGATGATGATGATGATGATGAT 420  
DB 633 TATGAGCATCGATAGTATCTGTGACCCAGTGCAGTGCAGAGAGAGGAGTCACTCAA 692

QY 458 CCAACCAGGAGCCTGCAGTGATAGCTGGACACAGGGACAAATTGCCCTGGAGTGCTGAACC 517

```

/lab_host="DH10B (phage-resistant)"
/Note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      287 a      366 c      259 g      247 t
ORIGIN
Query Match      40.1%; Score 410.4; DB 14; Length 1159;
Best Local Similarity 91.8%; Pred. No. 6e-112;
Matches 468; Conservative 0; Mismatches 36; Indels 5; Gaps 3;
QY 125 TTTTCTTCAAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGGTGTCTCA 184
Db 7 TTTTCTTCAAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGGTGTCTCA 66
QY 185 ATGCCGTCAGAGCTTGTCTCTGTGCTTTTGAAGGAGCGGAAAGCTGTGTACTC 244
Db 67 ATGCCGTCAGAGCTTGTCTCTGTGCTTTTGAAGGAGCGGAAAGCTGTGTACTC 126
QY 245 ACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTCCTG 304
Db 127 ACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTCCTG 186
QY 305 CTTATATTGTGTGTCGCCAGAGCTCCAGACTGTAAAAAATTCGAATACAGGCTAGC 364
Db 187 CTTATATTGTGTGTCGCCAGAGCTCCAGACTGTAAAAAATTCGAATACAGGCTAGC 246
QY 365 GAGCGTCAATTCTATCTACTGTGAACCTAGTATGATGATCCAGAGAAATTTGCAAAAAGAG 424
Db 247 GAGCGTCAATTCTATCTACTGTGAACCTAGTATGATGATCCAGAGAAATTTGCAAAAAGAG 306
QY 425 TTACAGAGAAACACAGGATCATGTGTACATCCCAACAGGAGCTGCAGTGATAGCTG 484
Db 307 TTACAGAGAAACACAGGATCATGTGTACATCCCAACAGGAGCTGCAGTGATAGCTG 366
QY 485 G-ACAAAGGACAAATTCCTGGAAGTGTGAACAGGTTTCCCTTTTGGTGGATGCA-CTGGT 542
Db 367 GCACAAGGGACACTTTCCTGCTGCAAGTGTGAACAGGTTTCCCTTTTGGTGGATGCACTGGT 426
QY 543 GTACCTCTAGTGTGAGGAGGAATCTTCTGCTG-----AATGCAATTACAGTTAGGCTC 598
Db 427 GTACCTCTAGTGTGAGGAGGAATCTTCTGCTGGAATAAGCCCTCTACAGTTAAGGCTC 486
QY 599 TGAAACCTAGTGTGAAGTATATGCTGCTG 628
Db 487 CTGACACCCCGCTGAAAGTATTATGCTG 516

RESULT 14
LOCUS      BQ885621
DEFINITION AGENCOURT_8754012 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332599
5', mRNA sequence.
ACCESSION BQ885621
VERSION   BQ885621.1 GI:22277639
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 918)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM13790 row: h column: 08  
High quality sequence stop: 609.  
Location/Qualifiers  
1..918  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6332599"  
/clone\_lib="NIH\_MGC\_130"  
/lab\_host="DH10B (phage-resistant)"  
/Note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."  
BASE COUNT 247 a 230 c 230 g 203 t 8 others  
ORIGIN  
Query Match 39.3%; Score 402.4; DB 14; Length 918;  
Best Local Similarity 83.7%; Pred. No. 1.4e-109;  
Matches 512; Conservative 0; Mismatches 92; Indels 8; Gaps 5;  
QY 1 ATGTGTGCTCAGTATGTGATCTCTCTTGTGTGATGTGAAAAAGCTCATATCAACATTCGA 60  
Db 308 ATGTGTGCTCAGTATGTGATCTCTCTTGTGTGATGTGAAAAAGCTCATATCAACATTCGA 367  
QY 61 GATTCTATCCACCTCACACCAGTGTACAAAGCTCCATTTTGAATCAACATCAAGGCGC 120  
Db 368 GACTCTATCCACCTCACCCAGTGTACAAAGCTCCATTTTGAATCAACATCAAGGCGC 427  
QY 121 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGTGCT 180  
Db 428 AATCTTTTCTTCAAAATGTGAACCTCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGTGCT 487  
QY 181 CTCAATGCGCTCAGAAAGCTTGTGTCTGTGATGCTTTAGAAAGAAAGCCGAAAGCTGTGT 240  
Db 488 CTTAATGCCATCAGAGGCTTAATCTCTGACGCCAGAGAGAAAGCCGAAAGCTGTGT 547  
QY 241 ACTCACAGCTGGAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300  
Db 548 ACTCACAGCTGGAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 607  
QY 301 CTGCTCTATATTTGGTGTGCCAGACAGCTCCAGAGCTGTAAAAAATTCGAATACAGCC 360  
Db 608 CTGCTCTATATTTGGTGTGCCAGACAGCTCCAGAGCTGTAAAAAATTCGAATACAGCC 667  
QY 361 TACGAGGCTCAATTTGATATCTGTGAACCTAGTATGATGATGATGATGATGATGATGATGAT 420  
Db 668 TATGGAGCATCATAGTATCTGTGACCCCAAGTGACGAGTCCAGAG-AAAAGTCACTCA 726  
QY 421 AGAGTTACAGAGAAACAGAGGATCATGTGTATCTCCCAACAGGAGCTGCAGTGATA 480  
Db 727 AGAATTTTATGAGAACAGAGGATCTTGGTGTGCCATCCCAACAGGAGCTGCAGTGATA 786  
QY 481 GCTGGACAAGGACAAATTCCTTGGAAAGTGTGGAACAGGTTTCTTGTG-TGGATCACT 539  
Db 787 GCTGGACNAGGAACATTCCTTGGAAAGTGTGGAACAGGTTTCTTGTG-TGGATCACT 846  
QY 540 -GGTGGTACCTTAGGT---GGAGGAGGAATGCTGTGGAATA---GCAATTACAGTTAA 593  
Db 847 GGTGGTACCTTAGGT---GGAGGAGGAATGCTGTGGAATA---GCAATTACAGTTAA 906  
QY 594 GCTCTGGAACC 605  
Db 907 GGCCCTGGAACC 918

RESULT 15  
LOCUS B1738966  
DEFINITION B1738966 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5367043 5',

DB 694 CAAAGATTAT-TGCAGAAACAGAGGCGATCTTTGGTCCATCCCAACCAGAGCCTGCAGTG 752  
QY 478 ATAGCTGGACAGGAGCAATTCGCCCTGGAAAGTGTGAACCAAGGTTCCCTTTG 528  
DB 753 ATAGCTGGACAGGAGCAATTCGCCCTGGAAAGTGTGAACCAAGGTTCCCTTTG 803

Search completed: June 24, 2003, 10:54:14  
Job time : 1468 secs

Accession BI738966  
Version BI738966.1 GI:15715979  
Keywords EST.  
Source house mouse.  
Organism Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 823)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1933 row: p column: 20  
High quality sequence stop: 799.  
Location/Qualifiers  
1..823  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5367043"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
237 a 208 c 193 g 184 t 1 others

# FEATURES

source

## BASE COUNT

ORIGIN

Query Match 38.3%; Score 391.8; DB 13; Length 823;  
Best Local Similarity 88.5%; Pred. No. 2e-106;  
Matches 470; Conservative 0; Mismatches 57; Indels 4; Gaps 4;  
QY 1 ATGTGCTCAGTATGTGATCTCTTTGCTGATGTTGAAAGCTCATATCAACATTGCA 60  
DB 274 ATGTGCTCAGTATGTGATCTCTTTGCTGATGTTGAAAGCTCATATCAACATTGCA 333  
QY 61 GATCTATCCACCTCACACAGTCTTACACAGTCCATTTTGAATCAACTACAGGCGC 120  
DB 334 GACTCTATCCACCTCACACAGTCTTACACAGTCCATTTTGAATCAACTACAGGCGC 393  
QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTTGAATTCGTGCTGT 180  
DB 394 AATCTTTTCTCAATGTGAATCTTCCAGAAACAGGATCTTTTGAATTCGTGCTGT 453  
QY 181 CTAAATGCGGTGAGAGTGTGCTGCTGATGTTAGAAAGGAGCCGAAAGCTGTTGTT 240  
DB 454 CTAAATGCGGTGAGAGTGTGCTGCTGATGTTAGAAAGGAGCCGAAAGCTGTTGTT 513  
QY 241 ACTACAGCAGTGGAAACCATG-GCCAGGCTCTACCTATGCTGCCAATTTGGAAGGAAT 299  
DB 514 ACTACAGCAGTGGAAACCATG-GCCAGGCTCTACCTATGCTGCCAATTTGGAAGGAAT 573  
QY 300 TCCTGCTTATATTGTTGGTCCCGCCACAGCTCCAGACTGTAA-AAACCTTGCATACAAAG 358  
DB 574 TCCTGCTTATATTGTTGGTCCCGCCACAGCTCCAGACTGTAA-AAACCTTGCATACAAAG 633  
QY 359 CCTACGAGCGCTCAATTTGATATAC-TGTGAACCTAGTGTGATGATCCAGAGAAATGTTGCA 417  
DB 634 CCTATGAGCATCGATAGTATATCTTGTGACCCAGTGACGAGTCCAGAGAAAGGTCAT 693  
QY 418 AAAAGAGTTACAGAGAAACAGAGGATCATGTTACATCCCAACAGGAGCCTCCAGTG 477

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:16:55 ; Search time 77 Seconds  
(without alignments)  
4074.420 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtgtcagattgcat.....atcagctgtttctgttttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1770	4	US-09-789-300A-1
2	1020	99.7	1020	4	US-09-789-300A-3
c 3	62.4	6.1	2299	4	US-09-153-599A-1
4	57	5.6	1545	2	US-08-628-039-1
5	57	5.6	1545	2	US-08-628-039-5
6	57	5.6	1545	2	US-08-628-039-7
7	57	5.6	1545	2	US-08-628-039-8
8	57	5.6	1545	2	US-08-673-388-1
9	57	5.6	1545	2	US-08-673-388-5
10	57	5.6	1545	2	US-08-673-388-7
11	57	5.6	1545	2	US-08-673-388-8
12	57	5.6	1545	2	US-08-614-877-1
13	57	5.6	1545	2	US-08-614-877-5
14	57	5.6	1545	2	US-08-614-877-7
15	57	5.6	1545	3	US-08-614-877-8
16	57	5.6	1545	3	US-08-912-205-1
17	57	5.6	1545	3	US-08-912-205-5
18	57	5.6	1545	3	US-08-912-205-7
19	57	5.6	1545	3	US-08-912-205-8
20	57	5.6	1545	4	US-09-440-400-1
21	57	5.6	1545	4	US-09-440-400-5
22	57	5.6	1545	4	US-09-440-400-7
23	57	5.6	1545	4	US-09-440-400-8
24	49.2	4.8	12720	1	US-08-403-866-11
c 25	46.8	4.6	7218	1	US-08-232-463-14
c 26	46	4.5	2698	4	US-08-936-165A-207
27	38.4	3.8	1485	4	US-09-088-435-2

28	36.6	3.6	6211	4	US-08-961-527-8	Sequence 8, Appli
29	36	3.5	1275	4	US-09-134-001C-1039	Sequence 1039, Ap
30	34.2	3.3	2085	2	US-08-283-917-8	Sequence 8, Appli
31	34.2	3.3	2085	2	US-08-961-716-8	Sequence 8, Appli
32	33.8	3.3	1547	4	US-09-356-818A-1	Sequence 1, Appli
33	33.8	3.3	1910	1	US-08-247-902A-1	Sequence 1, Appli
34	33.8	3.3	1910	5	PCT-US93-10541-1	Sequence 1, Appli
35	33.8	3.3	3257	5	PCT-US91-09784-1	Sequence 1, Appli
36	33.4	3.3	1274	4	US-09-227-357-75	Sequence 75, Appli
37	32.6	3.2	8494	4	US-08-961-527-163	Sequence 163, App
38	32	3.1	1557	4	US-09-134-001C-1131	Sequence 1131, App
39	32	3.1	3319	2	US-08-960-022-19	Sequence 19, Appli
40	32	3.1	4651	2	US-08-449-644-4	Sequence 4, Appli
41	32	3.1	4651	2	US-08-087-244A-4	Sequence 4, Appli
42	31.8	3.1	351	4	US-09-556-877-279	Sequence 279, App
43	31.8	3.1	351	4	US-09-620-412C-279	Sequence 279, App
44	31.6	3.1	289	4	US-09-007-005-17	Sequence 17, Appli
45	31.6	3.1	289	4	US-09-244-796-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1  
US-09-789-300A-1  
; Sequence 1, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1770  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (69)...(1091)  
US-09-789-300A-1

Query Match	100.0%	Score 1023;	DB 4;	Length 1770;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1023;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	ATGTGTGTCAGTATTCGATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA	60	
DB	69	ATGTGTGTCAGTATTCGATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA	128	
QY	61	GATTCCTATCCACTCACACAGTCTCAACAGCTCCATTTTGAATCAACTAACAGGCGC	120	
DB	129	GATTCCTATCCACTCACACAGTCTCAACAGCTCCATTTTGAATCAACTAACAGGCGC	188	
QY	121	AATCTTTTCTTCAAAATGTGAACCTTTCCAGAAAAAGGATCTTTTAAAGATTCGTGTGCT	180	
DB	189	AATCTTTTCTTCAAAATGTGAACCTTTCCAGAAAAAGGATCTTTTAAAGATTCGTGTGCT	248	
QY	181	CTCAATGCGCTCAGAAAGCTTGTTCTGATGCTTTTGAAGAGCGGAAAGCTGTGTT	240	
DB	249	CTCAATGCGCTCAGAAAGCTTGTTCTGATGCTTTTGAAGAGCGGAAAGCTGTGTT	308	
QY	241	ACTCACAGCAGTGGAAACCATGCCCAGGCTCTCACCTATGCTGCCAAATTTGGAGAAAT	300	
DB	309	ACTCACAGCAGTGGAAACCATGCCCAGGCTCTCACCTATGCTGCCAAATTTGGAGAAAT	368	
QY	301	CCTGCTTATATTGTGTGTCGCCCCAGACAGCTCCAGAACTGTAAAAAACTTGCATACAAGCC	360	



Db 369 CTTGCTTATATGTTGGTGGCCACAGAGCTCCAGACTGTAAAAAATTTGCAATACAGGCC 428  
Qy 361 TACGGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420  
Db 429 TACGGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 488  
Qy 421 AGATTACAGAAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCAGTGATA 480  
Db 489 AGATTACAGAAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCAGTGATA 548  
Qy 481 GCTGGCAAGGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 540  
Db 549 GCTGGCAAGGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 608  
Qy 541 GTGTAACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTCAGATTAAGGCTCTG 600  
Db 609 GTGTAACCTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTCAGATTAAGGCTCTG 668  
Qy 601 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAAATCAGATGACTGCTACCAAGTCC 660  
Db 669 AAACCTAGTGTGAAGGTATATGCTGTGAACCTCAAAATCAGATGACTGCTACCAAGTCC 728  
Qy 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 720  
Db 729 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 788  
Qy 721 AAATCCAGCATTTGGCTTGAACACCTGCTTATATCAGGAGACCTTTGTGGATGATATCTTC 780  
Db 789 AAATCCAGCATTTGGCTTGAACACCTGCTTATATCAGGAGACCTTTGTGGATGATATCTTC 848  
Qy 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGCTGTGGAGAGGATGAACACTA 840  
Db 849 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGCTGTGGAGAGGATGAACACTA 908  
Qy 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 900  
Db 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 968  
Qy 901 GTTTCCTCCAGAGTAAAGAACATTTGTTATGCTCAGTGTGGAATGTAGACTTAAACC 960  
Db 969 GTTTCCTCCAGAGTAAAGAACATTTGTTATGCTCAGTGTGGAATGTAGACTTAAACC 1028  
Qy 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1088  
Qy 1021 TAA 1023  
Db 1089 TAA 1091

## RESULT 2

US-09-789-300A-3  
; Sequence 3, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; TITLE OF INVENTION: Dependent Enzyme Family Member and Uses Therefore  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1020  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-789-300A-3

Query Match 99.7%; Score 1020; DB 4; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTGTCTCAGTATTCATCTCCTTTGGCTGTGTTGAAAGCTCATATCAACATTCGA 60  
Db 1 ATGTGTCTCAGTATTCATCTCCTTTGGCTGTGTTGAAAGCTCATATCAACATTCGA 60  
Qy 61 GATTCTATCCACCTCACACCAGTCTTAACAGCTCCATTTTGAATCAACTAACAGGGCG 120  
Db 61 GATTCTATCCACCTCACACCAGTCTTAACAGCTCCATTTTGAATCAACTAACAGGGCG 120  
Qy 121 AATCTTTTCTTCAAAATGTGAACCTCTCCAGAAACAGGATCTTTTAAGATTCGTGGTCT 180  
Db 121 AATCTTTTCTTCAAAATGTGAACCTCTCCAGAAACAGGATCTTTTAAGATTCGTGGTCT 180  
Qy 181 CTCATGCTCCCTCAGAAAGCTTTGGTTCCTGTGCTTTAGAAAGAGCCGAAAGCTGTTGT 240  
Db 181 CTCATGCTCCCTCAGAAAGCTTTGGTTCCTGTGCTTTAGAAAGAGCCGAAAGCTGTTGT 240  
Qy 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAAT 300  
Db 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAAT 300  
Qy 301 CCTGCTTATATTTGTGTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
Db 301 CCTGCTTATATTTGTGTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
Qy 361 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420  
Db 361 TAGCGAGCGTCAATTTGTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420  
Qy 421 AGAGTTACAGAAAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCAGTGATA 480  
Db 421 AGAGTTACAGAAAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCTGCAGTGATA 480  
Qy 481 GCTGGCAAGGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 540  
Db 481 GCTGGCAAGGGACAATTTGCCCTGGAGTGTCTGAACCAAGTTCCTTTGGTGGATGCACTG 540  
Qy 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTTGCTGGAATAGCAATTAAGTTAAGGCTCTG 600  
Db 541 GTGGTACCTGTAGTGGAGGAGGAATGCTTTGCTGGAATAGCAATTAAGTTAAGGCTCTG 600  
Qy 601 AARCTTGTGAGGTATATGCTGTGAACCTCAAAATGCAGATGACTGCTACCAGTCC 660  
Db 601 AARCTTGTGAGGTATATGCTGTGAACCTCAAAATGCAGATGACTGCTACCAGTCC 660  
Qy 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 720  
Db 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 720  
Qy 721 AATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGACCTTTGTGGATGATATCTTC 780  
Db 721 AATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGACCTTTGTGGATGATATCTTC 780  
Qy 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGCTGTCTCAACATTTTCAAACT 840  
Db 781 ACTGTACAGAGGATGAATTAAGTGTGAACCCAGCTGCTGTCTCAACATTTTCAAACT 840  
Qy 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 900  
Db 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 900  
Qy 901 GTTTCCTCCAGAGTAAAGAACATTTGTTATGCTCAGTGGTGGAAATGTAGACTTAAACC 960  
Db 901 GTTTCCTCCAGAGTAAAGAACATTTGTTATGCTCAGTGGTGGAAATGTAGACTTAAACC 960  
Qy 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Db 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020



## RESULT 3

US-09-153-599A-1/c  
; Sequence 1, Application US/09153599A  
; Patent No. 6420177  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Method for Strain Improvement of  
; TITLE OF INVENTION: Erythromycin Producing Bacterium  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 180 N. Stetson Avenue, 2 Prudential Plaza  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa V.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: FER2159P0041US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2299 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-153-599A-1

Query Match 6.1%; Score 62.4; DB 4; Length 2299;  
Best Local Similarity 49.5%; Pred. No. 4.1e-10;  
Matches 190; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

Qy	509	TGCTGAACCAAGTTCTTTGGTGATGCACCTGTGGTGTACTGTAGTGGAGGAAATGC	568
Db	2296	TCTGGAGCAGTTGCCGAGGTCGGACAGTGTGTGCCCGAGGCGGCGGGGTGG	2237
Qy	569	TTGCTGGAATAGCAATTACAGTTAAGGCTCTGAACCTAGTGTCAAGGTATATGCTGCTG	628
Db	2236	TCAGCGGCATCGCGCGCGCGTGAAGCGGAGCACCAGGTCGCGGTGCTCGCGGTC	2177
Qy	629	AACCTCAATGCAGATGACTGTACCAAGCTTCAAGCTGAAGGGGAAACTGATGCCCAATC	688
Db	2176	AAGCGGAGCAGCGCGCGCTGCGCGCGTGGCGGCGGGAACCGGTGCTCTGCTGC	2117
Qy	689	TTTATCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTG---GCTTGAACACCT	745
Db	2116	TCGACACCCAGCGGACGATGGCGGCATCGCGGTCCCGCGCGGAGCGAGCTGACCT	2057
Qy	746	GGCTATTATCAGGACCTTGTGGATCATATCTTCACTGTACAGAGGATGAATTAAGT	805
Db	2056	TCGCCCCACGTACGAGCTGTGCACACGTGTCTACGGTGGGAGGAGCGCTCTCGC	1997
Qy	806	GTGCAACCCAGCTGGTGTGGGAGGATGAACCTACTTCAATGAACCTACACAGCTGGTGTG	865
Db	1996	GCAGCTGTCTGTGTGCTGGAAGCGGCGAAGCTGGTGGTGCAGACCCCGCGGCTCGCG	1937
Qy	866	GAGTGGCTGTGCTGTCTCAAC	889

Db 1936 CGTCCGCCGACTGCTGGAGCAC 1913

## RESULT 4

US-08-628-039-1  
; Sequence 1, Application US/08628039  
; Patent No. 5942660  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchey, Maud A. W.  
; APPLICANT: Clemente, Thomas E.  
; APPLICANT: Connor-Ward, Danielle V.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
; TITLE OF INVENTION: Biosynthesis of poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
; TITLE OF INVENTION: in Bacteria and Plants  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB  
; STREET: 800 No. 5942660th Lindbergh Boulevard  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63167  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE: 13-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-1

Query Match 5.6%; Score 57; DB 2; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 1.9e-08;  
Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy	216	AGAAGGAAGCCGAAGCTGTGTACTACAGCAGTGGAAACCATGGCCAGGCTCTCAC	275
Db	225	AGAAGGAAGCCGACGCGTGTCTTCTGCGGTAAACCCAGCGGAGGCGTCCG	284
Qy	276	CTATGCTGCCAAATTTGAAGGAATTCCTGCTTATATTGTGGTGGCCCCAGACAGCTCAGA	335
Db	285	GTTTCTTCTGCGCGGTAGGCGGTGGAAGCGCTGATGCTTATGCCAACCCGCGCCGA	344
Qy	336	CTGTAAAAAATTGCAATACAAGCTACGGAGCGCTCAATTGTATACTGTGAACCTAGTGA	395

Db 345 CATCAAGTCAGCGGCTGCGCGCTTCGGCGGGAAGTGTCTCCAGCGCGGAACCT 404  
Qy 396 TGAGTCCAGAGAAATGTTGCAAAAAGAGTTACAGAGAAACAGAGGCATCATGGTACA 455  
Db 405 TGATGAAGCGAAACGCAAGCGATCGCAACTGTACACAGCAGGGTTACCTGGTGCC 464  
Qy 456 TCCCAACAGAGCGCTCGAGTGTAGCTGGAAGGACAAATTCCTTGAAGTGTCAA 515  
Db 465 GCCGTTCGACCATCCGATGTTGATTCGGGCAAGGACGCTGCGCTGGAAGTGTCAA 524  
Qy 516 CCAGGTTCTTTGGTGTGATGACATGGTGTACCTGTAGTGGAGAGAAATGCTTGTGG 575  
Db 525 CGAGGAGCCCATCGACCGCGATTTGTGCCAGTGGCGGCGCGGTCTGGCTGTG 584  
Qy 576 AATAGCAATTACAGTTAAGGCTGTAACCTGTAAGCTAGTGTGTAAGCTATATGCTGAACCTC 635  
Db 585 CGTGGCGGTGTGATCAAACTGATGCGCGCAATCAAGTGTACCGGTGAGAACGCGGA 644  
Qy 636 AATGCAAGTGTGCTACCATGCTCAAGCTGGAAGGGAACCTGATGCCAATCTTTATCC 695  
Db 645 AGACTCGCGCTGCTGAAAGCAGCGCTGGATGCGGGTCACTCGGTTGATCTGCGCGCGT 704  
Qy 696 TCCAGAAACCATACAGATGGTGTCAATCCAGATTTGGCTTGAACACCTGGCCCTATAT 755  
Db 705 AGGCTATTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCGTTTATG 764  
Qy 756 CAGGACCTTGTGATGATCTTCTACTGTACAGAGGATGAAAT 800  
Db 765 CCAGGATATCTGACGACATCATCACCGTCGATAGCGATGCGAT 809

## RESULT 5

US-08-628-039-5  
; Sequence 5, Application US/08628039  
; Patent No. 5942660

## GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.  
; APPLICANT: Connor-Ward, Thomas E.  
; APPLICANT: Fedele, Mary J.  
; APPLICANT: Fry, Joyce E.  
; APPLICANT: Howe, Arlene R.  
; APPLICANT: Rozman, Renee J.

; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate

; TITLE OF INVENTION: in Bacteria and Plants

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Gary M. Bond, Monsanto Company, A3SB

; STREET: 800 No. 5942660th Lindbergh Boulevard

; CITY: St. Louis

; STATE: Missouri

; COUNTRY: USA

; ZIP: 63167

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/628,039

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE: 13-MAR-1996

; ATTORNEY/AGENT INFORMATION:  
; NAME: Bond, Gary  
; REGISTRATION NUMBER: 29,283  
; REFERENCE/DOCKET NUMBER: 38-21(13585)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)694-3412  
; TELEFAX: (314)695-5435  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-628-039-5

Query Match 5.6%; Score 57; DB 2; Length 1545;  
Best Local Similarity 43.6%; Pred. No. 1.9e-08;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGCAAGCCGAAAGCTGTTCTACTACAGCAGTGGAAACCATGCCAGGCTCTCAC 275

Db 225 AGAACAAAGCCGACCGCTGATCTGTTGCGGGTAACCAACGCGGCGGTGCG 284

Qy 276 CTATGCTGCCAAATTTGAAGGAATTCCTGCTTATATTGTGTGCTCCACAGCAGCTCCAGA 335

Db 285 GTTTCTCTCGCGGTTAGCGGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344

Qy 336 CTGTAAAAAATTTGCAATACAGCCTACGAGCGTCAATTTGTATCTGTGAACCTAGTGA 395

Db 345 CATCAAAAGTCGACCGCTGCGCGGCTTCGGCGGGAAGTGTCTCCACGCGCGGAACCT 404

Qy 396 TGAGTCCAGAGAAATTTGCATAAAGAGTTACAGAGAAACAGAAAGGCATCATGGTACA 455

Db 405 TGATGAAGCAAGCAAGCGAAGCGATCGAACTGTTCACAGCAGCGGGTTACCTGGGTGCC 464

Qy 456 TCCCAACAGGAGCGCTGCAGTGTAGCTGCAGAGGAGCAATTCGCCCTGGAAGTGTCAA 515

Db 465 GCCGTTCCGACCATCCGATGTGATTGCGGGCAAGGACCGCTGGCGCTGGAAGTGTCCA 524

Qy 516 CCAGGTTCTTTGTTGGATGCATCTGAGTGTACCTGTAGTGTAGTGGAGGAGGAATGTTGTGG 575

Db 525 GCAGGAGCGCCCATCTCGACCGCGTATTTGTGCCAGTCCGGCGGCGGTCTGGCTGTG 584

Qy 576 AATAGCAATTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTCTGAACCTC 635

Db 585 CGTGGCGGTCTGATCAACAACCTGATGCGCGCAAAATCAAAAGTATCCCGCTGAAGCGGA 644

Qy 636 AATGCAAGTGTGCTTACCAGTCCAGCTGAAGGGGAAACCTGATGCCCAATCTTTATCC 695

Db 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATGCGGGTCACTCCGTTGATCTGCGCGCGT 704

Qy 696 TCCAGAAACCATACAGATGGTGTCAATCCAGATTTGGCTTGAACACCTGGGCTATAT 755

Db 705 AGGCGTATTTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCGTTTATG 764

Qy 756 CAGGAGCCTTGTGATGATATCTTCACTGTCAAGAGGATGAAT 800

Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGATGCGAT 809

## RESULT 6

US-08-628-039-7

; Sequence 7, Application US/08628039

; Patent No. 5942660

; GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.  
; APPLICANT: Mitsky, Timothy A.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stark, David M.  
; APPLICANT: Hinchee, Maud A. W.

576	ATATAGCAATTACAGTTAAGGCTCTGAACACCTAGTGTGAAGGTATATATGCTGTAACCCCTC	635
585	CGTGGCGGGTCTGATCAAAACAACCTAGTGGCGGCAAAATCAAAAGTATGCCCGTAGAAGCGGA	644
636	AAATGCGAGATGACTGTACCCAGTCCAAGCTGAAGGGGAAACTCATGCCCAATCTTTATCC	695
645	AGACTCGGCTGCTGAAAGCAGCGCTGGATCGGGGTCAATCCGGTTGATCTCGCGCGCT	704
696	TCCAGAAACCATAGCAGATGGTTCAAATCCAGCATTTGGCTTGAACACCTGGCTATTAT	755
705	AGGGCTATTTCCTGAAGCGGTAGCGGTAAACGCATCGGTGACGAAACCTTCCGTTTATG	764
756	CAGGGACCTTGTGGATGATATCTTCACTGTCCACAGAGGATGAAT	800
765	CCAGGAGTATCTCGACGACATCATCACCGTCGATACCGATCGCAT	809

RESULT 7  
 US-08-628-039-8  
 ; Sequence 8, Application US/08628039  
 ; Patent No. 5942660  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gruys, Kenneth J.  
 ; APPLICANT: Mitsky, Timothy A.  
 ; APPLICANT: Kishore, Ganesh M.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Padgett, Stephen R.  
 ; APPLICANT: Stark, David M.  
 ; APPLICANT: Hinchee, Maud A. W.  
 ; APPLICANT: Clemente, Thomas E.  
 ; APPLICANT: Connor-Ward, Dannette V.  
 ; APPLICANT: Fedele, Mary J.  
 ; APPLICANT: Fry, Joyce E.  
 ; APPLICANT: Howe, Arlene R.  
 ; APPLICANT: Rozman, Renee J.  
 ; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and  
 ; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval  
 ; TITLE OF INVENTION: In Bacteria and Plants  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Gary M. Bond, Monsanto Company, A35B  
 ; STREET: 800 No. 5942660th Lindbergh Boulevard  
 ; CITY: St. Louis  
 ; STATE: Missouri

```

? COUNTRY: US
? ZIP: 63167
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/628,039
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE: 13-MAR-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Bond, Gary
? REGISTRATION NUMBER: 29,283
? REFERENCE/DOCKET NUMBER: 38-21(13585)A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (314)694-3412
? TELEFAX: (314)695-5435
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1545 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-628-039-8

```

Query Match	5.6%;	Score 57;	DB 2;	Length 1545;
Best Local Similarity	43.0%;	Pred. No. 1.9e-08;		
Matches 255;	Conservative 0;	Mismatches 330;	Indels 0;	Gaps 0;

  

QY	216	AGAAGGAAGCCGAAAGCTCTGTTACTACAGCAGTGGAAACCATCGCCAGGCTCTCAC	275
Db	225	AGACAGAAGCCGACGGCGTGATCACTGTTCTCGGGTAACACCGCAGGCGTCCG	284
QY	276	CTATGTGCAAAATTTGAAGGAATTCCTGCTTATATTGTGTCGTCACAGACTCCAGA	335
Db	285	GTTTCTTCTTCGCGGTTAGCGCTGAAGGCCCTGATCGTTATGCCAACCCACCGCGGA	344
QY	336	CTGTAAAAAACTTGCATACAGCCCTACGAGCGTCAATTGTATCTGTGAACCTAGTGA	395
Db	345	CATCAAGTCGACCGGCTGCGCGGCTTCGCGCGGGAAGTCTGCTCCACCGCGCAACTT	404
QY	396	TGAGTCCAGAGAAAATCTTTGCAAAAAGAGTTACAGAAGAAACAGAGGCATCATGTACA	455
Db	405	TGATGAAGCGAAGCAAGCGATCGAATGTCACAGCAGCGGGTTCACTGGGTGCC	464
QY	456	TCCCAACACGAGCGCTGCAGTGTAGTGTGACAAAGGGACAATTGCCCTGGAAGTGTGAA	515
Db	465	GCGGTTTCGACCATTCGATCAAAACAATGATCGCGGCAAGGCACGCTGGCGTGAAGTCTCCA	524
QY	516	CAAGGTTCTCTTTGGTGGATGCACTGCTGGTGTACCTGTAGGTGAGGAGGAATGCTTGCTGG	575
Db	525	GCAGGACGCCCATCTCGACCGCGTATTGTGTCAGTCGCGCGCGCGGTCTGCTGCTTG	584
QY	576	AATAGCAATACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTC	635
Db	585	CGTGGCGGTGCTGATCAAAACAATGATCGCGCAAAATCAAAAGTATGATCGCCGTAGAAGCGGA	644
QY	636	AAATGCAGATGACTGTACCACTGTCAGCTCCAAGCTGAAGGGAAACTGATGCCCCAATCTTTATCC	695
Db	645	AGACTCCGCTGCTGAAAGCAGCGTGGATCGGGTCAATCGCGTTGATCTGCGCGCGCT	704
QY	696	TCAGAAACCATACAGCATGGTGTCAAACTCCAGCATTTGGCTTGAAACACCTTGGCTATTAT	755
Db	705	AGGCGTATTGTCTGAAGCGCTAGCGGTAAACGCATCGGTGACGAAACCTTCGCTTATG	764
QY	756	CAGGACCTGTGGATCATATCTTCACTGTCAACAGGAGGATGAAT	800
Db	765	CCAGGAGTATCTCGACGACATCATCCGTCGATAGCGATGCGAT	809

  

RESULT 8	
US-08-673-388-1	
: Sequence 1, Application US/08673388	
: Patent No. 5958745	
: GENERAL INFORMATION:	
: APPLICANT: Gruys, Kenneth J.	
: APPLICANT: Mitsky, Timothy A.	
: APPLICANT: Kishore, Ganesh M.	
: APPLICANT: Slater, Steven C.	
: APPLICANT: Padgett, Stephen R.	
: APPLICANT: Stark, David M.	
: APPLICANT: Hinchee, Maud A. W.	
: APPLICANT: Clemente, Thomas E.	
: APPLICANT: Connor-Ward, Dannette V.	
: APPLICANT: Fedele, Mary J.	
: APPLICANT: Fry, Joyce E.	
: APPLICANT: Howe, Arlene R.	
: APPLICANT: Rozman, Renee J.	
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and	
: TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalerate	
: TITLE OF INVENTION: in Bacteria and Plants	
: NUMBER OF SEQUENCES: 11	
: CORRESPONDENCE ADDRESS:	
: ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F	
: STREET: 700 Chesterfield Parkway No. 5958745th	
: CITY: St. Louis	
: STATE: Missouri	

## RESULT 9

US-08-673-388-5  
; Sequence 5, Application US/08673388  
; Patent No. 5958745

## GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth J.  
APPLICANT: Mitsky, Timothy A.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Padgett, Stephen R.  
APPLICANT: Stark, David M.  
APPLICANT: Hinchee, Maud A. W.  
APPLICANT: Clemente, Thomas E.  
APPLICANT: Connor-Ward, Dannette V.  
APPLICANT: Fedele, Mary J.  
APPLICANT: Fry, Joyce E.  
APPLICANT: Howe, Arlene R.  
APPLICANT: Rozman, Renee J.

TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera

TITLE OF INVENTION: in Bacteria and Plants

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

STREET: 700 Chesterfield Parkway No. 5958745th

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/673,388

FILING DATE: 13-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Cohen, Charles E.

REGISTRATION NUMBER: 34,565

REFERENCE/DOCKET NUMBER: 38-21(13618)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224

TELEFAX: (314)537-6047

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-673-388-5

Query Match 5.68; Score 57; DB 2; Length 1545;

Best Local Similarity 43.68; Pred. No. 1.9e-08;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGGAGCGGAAAGCTGTGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCAC 275

Db 225 AGACAGAAAGCGCACGGCTGATCACTGCTTCCGGGTAAACACCGCGGCGTCCG 284

Qy 276 CTATGCTGCCAAATGGAAGGAATTCCTGCTTATATTGTTGGTCCGCCAGACGCTCCAGA 335

Db 285 GTTTCTCTCGCGGGTTAGCGGTGAAGGCCCTGATCGTTATGCCACCGCCACGCCGA 344

Qy 336 CTGTAAAAACTTCAATACAGCTACGAGCGTCAATTGTATCTGTGAACCTAGTGA 395

Db 345 CATCAAAAGTACGCGGTGCGGGCTTCGGCGGCAAGTCTGCTCCACCGCGCGAACTT 404

Qy 396 TGAGTCCAGAGAAAATGTTCAAAAAGAGTTACAGAAGAAACAGAGGCATCATGGTACA 455

Db 405 TGATGAAGCGGAACGCAAGGATCGAAGTGTCTACACAGAGGGGTTACCTGGGTGCC 464

Qy 456 TCCCAACGAGGAGCCTGCAGTGATAGTGTGACAGAGGACAATGCCCTGGAAGTGCTGAA 515  
Db 465 GCCCTTCGACCATCCGATGGTGTATTCGCGGCAAGGCACGCTGCGCTGGAACCTCCA 524  
Qy 516 CCAGGTTCCCTTGGTGGATGCACCTGGTGGTACCTGTAGGTGGAGGAGGAATGCTTGGTGG 575  
Db 525 GCAGGACGCCCATCTCGACCGCGTATTGTGCCAGTGGCGGCGCGGCTCTGGCTGCTTG 584  
Qy 576 AATAGCAATTACAGTTAAGGCTCTGAAACCTTAGTGTGAAGGTATATGCTGCTGAACCTC 635  
Db 585 CGTGGCGGTCTGATCAACAACATGATGCCCAATCAAGTATCCCGCTAGAACGCGGA 644  
Qy 636 AATGTCAGATGACTGTACCAAGTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTATPC 695  
Db 645 AGACTCCGCTGCTGCTGAAAGCAGCGCTGGATGCGGCTCATCCGCTTGATCTGCCGCGCGT 704  
Qy 696 TCCAGAAACCATACACAGATGGTGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCCTATTAT 755  
Db 705 AGGCTATTGCTGAAGCGGTAGCGGTAAACGATCGGTGACGAAACCTTCCGTTTATG 764  
Qy 756 CAGGACCTTGTGGATGATATCTTCACTGTTCACAGAGGATGAAAT 800  
Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCCGATAGCGATGCGAT 809

## RESULT 10

US-08-673-388-7

; Sequence 7, Application US/08673388

; Patent No. 5958745

; GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth J.

APPLICANT: Mitsky, Timothy A.

APPLICANT: Kishore, Ganesh M.

APPLICANT: Slater, Steven C.

APPLICANT: Padgett, Stephen R.

APPLICANT: Stark, David M.

APPLICANT: Hinchee, Maud A. W.

APPLICANT: Clemente, Thomas E.

APPLICANT: Connor-Ward, Dannette V.

APPLICANT: Fedele, Mary J.

APPLICANT: Fry, Joyce E.

APPLICANT: Howe, Arlene R.

APPLICANT: Rozman, Renee J.

TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval

TITLE OF INVENTION: in Bacteria and Plants

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

STREET: 700 Chesterfield Parkway No. 5958745th

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/673,388

FILING DATE: 13-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Cohen, Charles E.

REGISTRATION NUMBER: 34,565

REFERENCE/DOCKET NUMBER: 38-21(13618)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224

TELEFAX: (314)537-6047

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 base pairs



Db 765 CCAGGACTATCTCGACGACATCATCACCGTCGATAGCGGAT 809

## RESULT 12

US-08-614-877-1

; Sequence 1, Application US/08614877

; Patent No. 5959179

; GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.

; APPLICANT: Mitsky, Timothy A.

; APPLICANT: Kishore, Ganesh M.

; APPLICANT: Slater, Steven C.

; APPLICANT: Padgett, Stephen R.

; APPLICANT: Stark, David M.

; APPLICANT: Hinchee, Maud A. W.

; APPLICANT: Clemente, Thomas E.

; APPLICANT: Connor-Ward, Dannette V.

; APPLICANT: Fedele, Mary J.

; APPLICANT: Fry, Joyce E.

; APPLICANT: Howe, Arlene R.

; APPLICANT: Rozman, Renee J.

; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera

; TITLE OF INVENTION: in Bacteria and Plants

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

; STREET: 700 Chesterfield Parkway No. 5959179th

; CITY: St. Louis

; STATE: Missouri

; COUNTRY: USA

; ZIP: 63198

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/614,877

; FILING DATE: 13-MAR-1996

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Cohen, Charles E.

; REGISTRATION NUMBER: 34,565

; REFERENCE/DOCKET NUMBER: 38-21(10695)A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (314)537-6224

; TELEFAX: (314)537-6047

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1545 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-614-877-1

Query Match 5.6% Score 57; DB 2: Length 1545;

Best Local Similarity 43.6% Pred.No. 1.9e-08;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Qy 216 AGAAGGAGCCGAAAGCTGTGTACTACAGCAGTGGAAACATGGCCAGGCTCTCAC 275

Db 225 AGAACAGAAAGCGACGCGGTGATCATCTGCTCGGGTAAACACCGCGAGGCGTCG 284

Qy 276 CTATGTCGCAATTTGAAGAAATTCCTGCTTATATTTGGTGGTCCCCAGACAGCTCAGA 335

Db 285 GTTTTCTTCTCGCGGTTAGGCGTGAAGGCCCTTGATGCGCAACCCGCCCGCGA 344

Qy 336 CTGTAAAAAATTCGATACAGCCCTACGAGCGCTCAATTTGTATATGTAACCTTAGTGA 395

Db 345 CATCAAAATCGACCGGTGCGCGGCTTCGCGCGGAGGATGCTGCTCCACGGCGCGA 404

Qy 396 TGAGTCCAGAGAAAATGTTGCAAAAAAGATTACAGAAAGAAACAGAGGCATCATGGTACA 455

Db 405 TGATGAAGCGAAACGCAACCGATCGAACTGTACACAGCAGCGGGTTCCACTGGGTGCC 464

Qy 456 TCCCAACAGGAGGCTGCACTGATGATGACAGAGGACAAATGCCCCCTGGAAGTGTGAA 515

Db 465 GCGGTTTCGACCATCCGATGATGATGTCGCGGCAAGGACGCGCTGCGGCTGGAACGTCTCCA 524

Qy 516 CCAGGTTCCCTTTGGTGGATGCACCTGGGTGCTAGCTGTAGGTGGAGGAGGAATGCTTGCTGG 575

Db 525 GCAGGAGCGCCATCTCGACCGCGTATTGTGCGCAGTCGCGCGCGCGGTCTGGCTGCTTG 584

Qy 576 AATAGCAATTTACAGTTTAAGGCTCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTC 635

Db 585 CGTGGCGGTGCTGATCAAAACAACTGATGCCGCCAAATCAAAGTGATGCCCGTGAAGCGGA 644

Qy 636 AATGCGAGATGACTGCTACCAAGCTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTATCC 695

Db 645 AGACTCCGCTGCTGCTGAAAGCAGCGGTGATGCGGGTCAATCCGGTTGATCTGCGCGCGCT 704

Qy 696 TCCAGAAACCATAGCAGATGCTGCAAAATCCAGCATTTGGCTTGAACACCTGGCCTATTAT 755

Db 705 AGGCTATTGCTGAAGCGGTAGCGGTAAACCATCATCCGTCGATAGCGATGCGAT 764

Qy 756 CAGGAGCCTTGTGGATGATATCTTCACTGTACAGAGGATGAAT 800

Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGATGCGAT 809

## RESULT 13

US-08-614-877-5

; Sequence 5, Application US/08614877

; Patent No. 5959179

; GENERAL INFORMATION:

; APPLICANT: Gruys, Kenneth J.

; APPLICANT: Mitsky, Timothy A.

; APPLICANT: Kishore, Ganesh M.

; APPLICANT: Slater, Steven C.

; APPLICANT: Padgett, Stephen R.

; APPLICANT: Stark, David M.

; APPLICANT: Hinchee, Maud A. W.

; APPLICANT: Clemente, Thomas E.

; APPLICANT: Connor-Ward, Dannette V.

; APPLICANT: Fedele, Mary J.

; APPLICANT: Fry, Joyce E.

; APPLICANT: Howe, Arlene R.

; APPLICANT: Rozman, Renee J.

; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

; TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval

; TITLE OF INVENTION: in Bacteria and Plants

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F

; STREET: 700 Chesterfield Parkway No. 5959179th

; CITY: St. Louis

; STATE: Missouri

; COUNTRY: USA

; ZIP: 63198

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/614,877

; FILING DATE: 13-MAR-1996

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Cohen, Charles E.

; REGISTRATION NUMBER: 34,565

; REFERENCE/DOCKET NUMBER: 38-21(10695)A

; TELECOMMUNICATION INFORMATION:

```

: TELEPHONE: (314)537-6224
: TELEFAX: (314)537-6047
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1545 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-614-877-5

```

Query Match	5.6%	Score 57:	DB 2:	Length 1545;
Best Local Similarity	43.6%	Pred. No. 1.9e-08;		
Matches 255;	Conservative 0;	Mismatches 330;	Indels 0;	Gaps 0;
QY	216	AGAAAGGAGCCGAAAGCTGTGTACTACAGCAGCTGGAAACCATGGCCAGGCTCTCAC	275	
Db	225	AGAACGAAGCCGACCGCGTGATCACTGCTTTCGGGGTAACCAACGCGAGGGCGTGC	284	
QY	276	CTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATTGTGTGCCCCAGACAGCTCCAGA	335	
Db	285	GTWTTCTCTCGCGGTTAGSGGTGAAGGCCCTGATGTTATGCCACCCGCCACGCCGA	344	
QY	336	CTGTAAAAACTTGCRAATACAAAGCCTACGAGCGTCAATTGTATCTGTGAACCTAGTGA	395	
Db	345	CATCAAAGTCGACCGCGCTGCGCGGCTTCGCGCGGAAAGTGCTGCTCCACGGCGGAAC	404	
QY	396	TGAGTCCAGAGAAAATGTTTCCAAAAGAGTTTACAGAAGAAACAGAGGCATCATGGTACA	455	
Db	405	TGATGAAGCGAAACGCNAAGCGATCGNACTGTCACAGCAGCGGGTTACCTGGGTGCC	464	
QY	456	TCCCAACACGAGGACGCTGCAGTAGCTGACAAAGGACAAATTCCTCGGAAGTCTGAA	515	
Db	465	GCGGTTTCGACCATCCGATGTTGATTCGCGGGCAAGGCACGTGCGCTGGAAC	524	
QY	516	CCAGGTTCTTTTGGTGGATGCACTGGTGGTACCTGTAGGTGGAGGAGGAATGCTTGCTG	575	
Db	525	GCAGGAGGCCATCTCGACCGGTATTGTGCGACGCGCGCGCGGTCTGGCTGTTG	584	
QY	576	AATAGCAATTACAGTTAAGGCTCTGAACCTAGTGTGAAGGTATATGCTGCTGAACCC	635	
Db	585	CGTGGCGGTGCTGATCAAAACACTGTCGCGCAAAATCAAAGTGATCGCCGTAGAAGCG	644	
QY	636	AAATGCAGATGACTGCTACCAAGTCCAAAGCTGAAGGGGAAACTGATGCCCAATCTTTAT	695	
Db	645	AGACTCCGCGTCCCTGAAAGCAGCGCTGGATCGGGGTCACTCCGGTTGATCTGCCGCG	704	
QY	696	TCCAGAAACCATAGCAGATGGTTCAAATCCAGCATTTGGCTTGAACACCTTGGCTATT	755	
Db	705	AGGGCTATTGCTGAAGGCGTACGGGTAAACGCATCGGTGACGAACCTTCGGCTTATG	764	
QY	756	CAGGGACCTTGTGGATGATATCTTCACTGTACAGAGGATGAAT	800	
Db	765	CCAGAGTATCTCGACGACATCATCAACCGTTCGATAGCGATGCGAT	809	

```

RESULT 14
US-08-614-877-7
: Sequence 7, Application US/08614877
: Patent No. 5959179
: GENERAL INFORMATION:
: APPLICANT: Gruys, Kenneth J.
: APPLICANT: Mitsky, Timothy A.
: APPLICANT: Klushore, Ganesh M.
: APPLICANT: Slater, Steven C.
: APPLICANT: Padgett, Stephen R.
: APPLICANT: Stark, David M.
: APPLICANT: Hincee, Maud A. W.
: APPLICANT: Clemente, Thomas E.
: APPLICANT: Connor-Ward, Dannel
: APPLICANT: Fedele, Mary J.
: APPLICANT: Fry, Joyce E.
: APPLICANT: Howe, Arlene R.

```



QY 696 TCCAGAACCATACGACATGGTCAAAATCCAGCATTGGCTTGAACACCTGGCCATTAT 755  
Db 705 AGGCTATTGGCTGAAGCGGTAGCGGTAAACACGATCGGTGACGAAACCTTCGGTTTATG 764  
QY 756 CAGGACCTTGTGATGATATCTTCACTGTGCACAGAGGATGAAAT 800  
Db 765 CCAGGACTATCTGACGACATCATCACCGTCGATAGCGATGCGAT 809

## RESULT 15

US-08-614-877-8

Sequence 8, Application US/08614877

Patent No. 5959179

GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth J.

APPLICANT: Mitsky, Timothy A.

APPLICANT: Kishore, Ganesh M.

APPLICANT: Slater, Steven C.

APPLICANT: Padgett, Stephen R.

APPLICANT: Stark, David M.

APPLICANT: Hinchey, Maud A. W.

APPLICANT: Clemente, Thomas E.

APPLICANT: Connor-Ward, Dannette V.

APPLICANT: Fedele, Mary J.

APPLICANT: Fry, Joyce E.

APPLICANT: Howe, Arlene R.

APPLICANT: Rozman, Renee J.

TITLE OF INVENTION: Methods of Optimizing Substrate Pools and

TITLE OF INVENTION: Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyvalera

TITLE OF INVENTION: in Bacteria and Plants

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Charles E. Cohen, Monsanto Company, B4F

STREET: 700 Chesterfield Parkway No. 5959179th

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/614,877

FILING DATE: 13-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Cohen, Charles E.

REGISTRATION NUMBER: 34,565

REFERENCE/DOCKET NUMBER: 38-21(10695)A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (314)537-6224

TELEFAX: (314)537-6047

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1545 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-614-877-8

## Query Match

Best Local Similarity 5.6%; Score 57; DB 2; Length 1545;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

Pred. No. 1.9e-08;

QY 216 AGAAGAACCGGAAACCTGTGTTACTACAGCAGTGGAAACCATGGCCAGGCTCTCAC 275

Db 225 AGAACAGAACGCGACGCGGTGATCACTGCTTTCGGGGTAACACGCGCAGGCGGCTCGC 284

QY 276 CTATGCTGCCAAATGGAAGAAATTCCTGCTTATATGTCGCCCCAGACAGCTCCAGA 335

Db 285 GTTTTCTTCTGCGCGGTAGCGGTGAAGGCCCTGATCGTTATGCCAACCCGCCACCGCGCA 344  
QY 336 CTGTAAAAAATTGCAATACAAAGCCTACGGAGCGCTCAATTGTATATCTGTGAACCTAGTGA 395  
Db 345 CATCAAAAGTCGACCGGCTTCGCGGCTTCGCGGCGGAAGTGTCTCCACGCGCGAACTT 404  
QY 396 TGAGTCCAGAGAGAAATGTTGCAAAAAGATTACAGAGAAACAGAGGCGATCATGTTGTA 455  
Db 405 TGATGAAGCGAAACGCAAGCGATCGAACTGTACAGCAGCAGGCGGTTCACTGGGTGCC 464  
QY 456 TCCCAACCCAGGAGCCTCGACGTAGCTGGACAAAGGACAAATTCGCCCTGGAAGTGTGAA 515  
Db 465 GCCGTTCCAGCCATCCGATGGTGTGTCGCGGCAAGGACCGCTGGCGCTGGAACCTGCTCCA 524  
QY 516 CCAGGTTCCCTTGGTGGATGCACCTGGTGGTACCTGTAGTGGAGGAGGAATGCTGTGCTGG 575  
Db 525 GCAGGACGCGCATCTCGACCGCGTATTTGTGCCAGTCGCGCGGCGGCTCTGGGCTGCTTG 584  
QY 576 AATAGCAATTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTC 635  
Db 585 CGTGGCGGTGCTGATCAAAACAACTGATGCGCGCAAAATCAAAAGTGATCGCCCTAGAAGCGGA 644  
QY 636 AAATGCGAGATGACTGTCTACCAGTCCAAGCTGAAGGGGAAACTGATGCCCAATCTTATCC 695  
Db 645 AGACTCCGCTGCTGAAAGCAGCGCTGGATGCGGGTCATCCGGTTGATCTGCGCGCGGT 704  
QY 696 TCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTGAACACCTGGCCCTATTAT 755  
Db 705 AGGCTATTGCTCAAGCGCTAGCGGTAAACCGCATCGGTGACCAAACTTCCGTTTATG 764  
QY 756 CAGGACCTTGTGGATGATATCTTCACTGTGCACAGAGGATGAAAT 800  
Db 765 CCAGGAGTATCTCGACGACATCATCACCGTCGATAGCGATGCGAT 809

Search completed: June 24, 2003, 06:24:42

Job time : 81 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:27:09 ; Search time 2668 seconds  
(without alignments)  
11104.444 Million cell updates/sec

Title: US-09-889-609B-1  
Perfect score: 1018  
Sequence: 1 atgtgtgtcagctactgcatt.....ttaccagagctgtctgttttaa 1018

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgt\_mus.\*
- 34: em\_hgt\_pln.\*
- 35: em\_hgt\_rod.\*
- 36: em\_hgt\_mam.\*
- 37: em\_hgt\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hgtgo\_hum.\*
- 40: em\_hgtgo\_mus.\*
- 41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1011.2	99.3	1197	10	AF148321	AF148321 Mus muscu
2	1011.2	99.3	1411	10	BC011164	BC011164 Mus muscu
3	807.8	79.4	1020	6	AX224420	AX224420 Sequence
4	807.8	79.4	1023	9	AY034081	AY034081 Homo sapi
5	807.8	79.4	1166	9	AF169974	AF169974 Homo sapi
6	807.8	79.4	1770	6	AX224418	AX224418 Sequence
7	807.8	79.4	2477	9	AK023169	AK023169 Homo sapi
8	806.2	79.2	1336	6	AX259575	AX259575 Sequence
9	664.6	65.3	2264	9	HS0805450	AL834378 Homo sapi
10	257.2	25.3	184151	10	AL604066	AL604066 Mouse DNA
11	257.2	25.3	205910	2	AC044847	AC044847 Mus muscu
12	250.8	24.6	3379	10	BC031531	BC031531 Mus muscu
13	212.2	20.8	101814	2	AC122087	AC122087 Rattus no
14	181.8	17.9	3807	9	BC019090	BC019090 Homo sapi
15	181.8	17.9	4140	9	AK026565	AK026565 Homo sapi
16	181.8	17.9	123454	9	HSBC17A99	AL450226 Homo sapi
17	181.8	17.9	221909	2	AC021705	AC021705 Homo sapi
18	151.4	14.9	40005	8	SPCC320	AL022245 S.pombe c
19	151.4	14.9	42689	8	SPCC330	AL031603 S.pombe c
20	142	13.9	295350	1	AP004826	AP004826 Staphyloc
21	140.2	13.8	4107	9	AB037822	AB037822 Homo sapi
22	130	12.8	6525	8	SCYKL218C	228218 S.cerevisia
23	130	12.8	24577	8	SC60RF	X75951 S.cerevisia
24	124.4	12.2	303750	1	AP003133	AP003133 Staphyloc
25	124.4	12.2	346900	1	AP003362	AP003362 Staphyloc
26	119.4	11.7	166050	1	AL646085	AL646085 Ralstonia
27	117.8	11.6	157609	2	AC095280	AC095280 Rattus no
28	116.8	11.5	16931	1	AE001716	AE001716 Thermotoga
29	114.4	11.2	298050	1	AP003189	AP003189 Clostridi
30	110.4	10.8	14449	1	AE004696	AE004696 Pseudomon
31	110.2	10.8	291050	1	AP000982	AP000982 Sulfolobu
32	108.6	10.7	35437	1	SCF43A	AL096837 Streptomy
33	107.4	10.6	2186	1	ECTDC	M21312 E.coli tdc
34	107.4	10.6	6295	1	ECTDCRAB	X14430 E. coli tdc
35	107.4	10.6	10516	1	AE000393	AE000393 Escherich
36	107.4	10.6	13979	1	AE005540	AE005540 Escherich
37	107.4	10.6	110000	1	ECOUW67_0	U18997 Escherichia
38	107.4	10.6	307962	1	AP002564	AP002564 Escherich
39	101	9.9	20878	1	AE008849	AE008849 Salmonell
40	97.8	9.6	258050	1	AL627278	AL627278 Salmonell
41	97.4	9.6	137	11	G32814	G32814 A009S27 Hum
42	97	9.5	303040	1	AP000991	AP000991 Thermopla
43	95.4	9.4	268200	1	AP000061	AP000061 Aeropyrum
44	94.8	9.3	10163	1	AE006661	AE006661 Sulfolobu
45	93.4	9.2	9990	1	AE010646	AE010646 Fusobacte

ALIGNMENTS

RESULT 1	AF148321	Mus musculus	serine racemase mRNA, complete cds.	1197 bp	linear	ROD 18-NOV-1999
LOCUS	AF148321	Mus musculus	serine racemase mRNA, complete cds.			
DEFINITION	AF148321	Mus musculus	serine racemase mRNA, complete cds.			
ACCESSION	AF148321	Mus musculus	serine racemase mRNA, complete cds.			
VERSION	AF148321.1	GI:6448864				
KEYWORDS						
SOURCE		Mus musculus				
ORGANISM		Mus musculus				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
TITLE		1 (bases 1 to 1197)				
		Wolosker, H., Blackshaw, S. and Snyder, S.H.				
		Serine racemase: a glial enzyme synthesizing D-serine to regulate				
		glutamate-N-methyl-D-aspartate neurotransmission				

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13409-13414 (1999)  
MEDLINE 20027561  
PUBMED 10557334  
REFERENCE 2 (bases 1 to 1197)  
AUTHORS Wolosker, H.  
TITLE Direct Submission  
JOURNAL Submitted (03-MAY-1999) Neuroscience, Johns Hopkins University, 725  
North Wolfe St., WBSB 806, Baltimore, MD 21205, USA

FEATURES  
source Location/Qualifiers  
1..1197  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
66..1085  
/EC\_number="5.1.1.17"  
/function="catalyzes L-serine to D-serine racemization"  
/codon\_start=1  
/product="serine racemase"  
/protein\_id="AAF08701.1"  
/db\_xref="GI:6448865"  
/translation="MCAOYCIISFADVEKAHINQDSIHLPVLTSSILNQIAGRNLF  
KCELPKTSFKIRGALNAIRGLIPDTPEEKPAVYTHSSGNHGQALTYAAKLEGIPA  
YIVVPTAENCKALQAGASIVCDPSDESKVYQRMQETEGILVHPNOEPVAVI  
AGQGTALVNLQVLDALVVPVGGGVAGIATIKALKPSVKVYAAEPNADDCY  
QSKLKGLEPNLHPETIADGVKSSIGLNTWPIIRLDVDDVFTVTEDEIKYATQLVWG  
RMLLLEPTAGVALAAVLVSOHFQTVSPVKVNCIVLSSGNDVLTSLNWMVQAEPA  
QTVSV"

BASE COUNT 345 a 283 c 283 g 286 t  
ORIGIN

Query Match 99.3%; Score 1011.2; DB 10; Length 1197;  
Best Local Similarity 99.7%; Pred. No. 1.9e-267;  
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTGTCTCAGTACGATCTCTTGTGTGTGTGTTGAAAGAGCTCATATCAACATTCAA 60  
Db 66 ATGTGTCTCAGTACGATCTCTTGTGTGTGTGTTGAAAGAGCTCATATCAACATTCAA 125  
QY 61 GACTCTATCACCTCACCCAGTCTTCAAGCTCCATTTTGAATCAATAGCAGGCGC 120  
Db 126 GACTCTATCACCTCACCCAGTCTTCAAGCTCCATTTTGAATCAATAGCAGGCGC 185  
QY 121 AATCTTTTCTCAATATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTGAGGTGCC 180  
Db 186 AATCTTTTCTCAATATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAAGATTGAGGTGCC 245  
QY 181 CTTAATGGCATCAGAGCGCTTAATCTGTGACACCCAGAGAGAGCCCAAGCCGTAGTT 240  
Db 246 CTTAATGGCATCAGAGCGCTTAATCTGTGACACCCAGAGAGAGCCCAAGCCGTAGTT 305  
QY 241 ACTCACAGCAGCGGAACCATGGCCAAAGCTCTCACCTATGCTCTAACTGGAAGGAATT 300  
Db 306 ACTCACAGCAGCGGAACCATGGCCAAAGCTCTCACCTATGCTCTAACTGGAAGGAATT 365  
QY 301 CTGCTTTACATTTGTTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCAAAGCC 360  
Db 366 CTGCTTTACATTTGTTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCAAAGCC 425  
QY 361 TATGGAGCATCATAGTATATCTGTGACCCAAAGTGTGACGAGTCCAGAGAAAGTCACTCAA 420  
Db 426 TATGGAGCATCATAGTATATCTGTGACCCAAAGTGTGACGAGTCCAGAGAAAGTCACTCAA 485  
QY 421 AGAATTTATGCAAGAAACAGAGCATCTTGGTCCATCCCAACAGGAGCGCTGCAGTGATA 480  
Db 486 AGAATTTATGCAAGAAACAGAGCATCTTGGTCCATCCCAACAGGAGCGCTGCAGTGATA 545  
QY 481 GCTGGCAAGGAACAATTCCTTGGAGTCTGCTGAAGTCTGCAACAGGTTCCTTGGTAGTCACTG 540  
Db 546 GCTGGCAAGGAACAATTCCTTGGAGTCTGCTGAAGTCTGCAACAGGTTCCTTGGTAGTCACTG 605  
QY 541 GTGGTACCAGTAGGAGGAGGAGGAATGGTGTGTTGGAATAGCCATTACATTAAGGCCCTG 600  
Db 606 GTGGTACCAGTAGGAGGAGGAGGAATGGTGTGTTGGAATAGCCATTACATTAAGGCCCTG 665

QY 601 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCACTCT 660  
Db 666 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCACTCT 725  
QY 661 AAACCTAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCACTAGCAGATGCTGTC 720  
Db 726 AAACCTAAAGAGAACTGACCCCAATCTTCATCTCCAGAAACCACTAGCAGATGCTGTC 785  
QY 721 AAATCCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGATCTTTC 780  
Db 786 AAATCCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGATCTTTC 845  
QY 781 ACTGTCAACGAGATGAAATCAAGTATGCAACCCAGCTGTGTGGGGGAGAGTGAACCTG 840  
Db 846 ACTGTCAACGAGATGAAATCAAGTATGCAACCCAGCTGTGTGGGGGAGAGTGAACCTG 905  
QY 841 CTCATTGAGCCGACTGCTGGCGTGGCAGTGGCTGCTGCTGCTCAGCATTTTCCAAACA 900  
Db 906 CTCATTGAGCCGACTGCTGGCGTGGCAGTGGCTGCTGCTCAGCATTTTCCAAACA 965  
QY 901 GTCTCTCCAGAGTAAAGAACGCTCTCCATGTACTCAGTGGGGGAGTGTAGACCTAAC 960  
Db 966 GTCTCTCCAGAGTAAAGAACGCTCTCCATGTACTCAGTGGGGGAGTGTAGACCTAAC 1025  
QY 961 TCCTTGAACCTGGTGGGCGAGCTGAACGCCAGCTCTTACAGACGCTGCTGTTT 1016  
Db 1026 TCCTTGAACCTGGTGGGCGAGCTGAACGCCAGCTCTTACAGACGCTGCTGTTT 1081

BC011164 1411 bp mRNA linear ROD 07-AUG-2002  
Mus musculus, Similar to serine racemase, clone MGC:18670  
IMAGE:4195695, mRNA, complete cds.  
BC011164  
BC011164.1 GI:15029880  
MGC.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1411)  
Strausberg, R.  
Direct Submission  
Submitted (25-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK plate: 24 Row: m Column: 6  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7305520.

FEATURES  
Location/Qualifiers  
1..1411  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

```
/map="FVB/N"
/clone="MGC:18670 IMAGE:4195695"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_Li9"
/lab_host="DH10B"
/notes="Vector: pCMV-Sport6"
288..1307
CDS
/codon_start=1
/product="Similar to serine racemase"
/protein_id="AAH11164.1"
/db_xref="GI:15029881"
/translacion="MCAQVCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLF
KELPKQSGFKIRGALNAIRGLIPDTPEEKRAVYTHSSGNHGQALTAAKLEGIPA
YIVVPTANCKRLAIQYASIVYCDPSDESKVTKRIMQETGILVHPNOEPYAVI
AQGTTALVNLQVPLVDALVVPVGGGVMAGTATIKALPKSVKYVAAEPNADDCY
QSKLKGELTPNLHPPTIADGVKSSISGLNTWPIIRDLVDVDFVTDEIKYATOLVWG
RMLLLEPTAGVALAAVLQSHQFTVSPVKVNCIVLSSGNGVDLTLNWWGQERPARPY
QTVSV"
BASE COUNT 411 a 335 c 340 g 325 t
ORIGIN
Query Match 99.3%; Score 1011.2; DB 10; Length 1411;
Best Local Similarity 99.7%; Pred. No. 1.9e-267;
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTGTGCTCAGTACGATCTCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCAA 60
Db 288 ATGTGTGCTCAGTACGATCTCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCAA 347
QY 61 GACTCTATCACTACCCAGTGTAAAGCTCCATTTTGAATCAATAGCAGGCGC 120
Db 348 GACTCTATCACTACCCAGTGTAAAGCTCCATTTTGAATCAATAGCAGGCGC 407
QY 121 AATCTTTTCTTCAATATGTAGCTCTTCCAGAAATCGGTCTTTTAAAGATTCAGGTGCC 180
Db 408 AATCTTTTCTTCAATATGTAGCTCTTCCAGAAATCGGTCTTTTAAAGATTCAGGTGCC 467
QY 181 CTTAATGCCATCAGAGGCTTAATCTGTACGCCAGAGAGAGAGGCCAAAGCCGTAGTT 240
Db 468 CTTAATGCCATCAGAGGCTTAATCTGTACGCCAGAGAGAGAGGCCAAAGCCGTAGTT 527
QY 241 ACTCAGCAGCGGGAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGAGGAAT 300
Db 528 ACTCAGCAGCGGGAACCATGCGCAAGCTCTCACCTATGCTGCTAACTGGAGGAAT 587
QY 301 CTGCTTACATTTGGTTCCTCCAAACAGCTTCCCAACTGCAAGAACTGGCAATCCAGCC 360
Db 588 CTGCTTACATTTGGTTCCTCCAAACAGCTTCCCAACTGCAAGAACTGGCAATCCAGCC 647
QY 361 TATGGAGCATCGATAGTATCTGTGACCCAAAGTGACGAGTCCAGAGAAAAGGTCACTCAA 420
Db 648 TATGGAGCATCGATAGTATCTGTGACCCAAAGTGACGAGTCCAGAGAAAAGGTCACTCAA 707
QY 421 AGAATTTATCAAGAAACAGAGGCATCTTGGTCCATCCCAACAGGAGCTGCAAGTATA 480
Db 708 AGAATTTATCAAGAAACAGAGGCATCTTGGTCCATCCCAACAGGAGCTGCAAGTATA 767
QY 481 GCTGGACAAGGACAAATGCCCTGGAAAGTCTGAAACAGGTTCCCTTTGGTAGATGCACTG 540
Db 768 GCTGGACAAGGACAAATGCCCTGGAAAGTCTGAAACAGGTTCCCTTTGGTAGATGCACTG 827
QY 541 GTGGTACCAGTAGGAGGAGGAATGGTTGTGTGGAAATAGCCATTAACAATTAAGGCCCTG 600
Db 828 GTGGTACCAGTAGGAGGAGGAATGGTTGTGTGGAAATAGCCATTAACAATTAAGGCCCTG 887
QY 601 AAACCTAGTCTGAAGTATACGCTGTGAGCCCTCAATGCCAGATCACTGCTACCACTCT 660
Db 888 AAACCTAGTCTGAAGTATACGCTGTGAGCCCTCGAATGCCAGATCACTGCTACCACTCT 947
QY 661 AAACCTGAAAGGAGAACTGACCCCAATCTTCACTCTCCAGAAACCATAGCAGATGTGTC 720
Db 948 AAACCTGAAAGGAGAACTGACCCCAATCTTCACTCTCCAGAAACCATAGCAGATGTGTC 1007
```

```
721 AAATCCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTTGTGGATGATGCTTC 780
1008 AAATCCAGCATTTGGCTTGAATACCTGGCCTATTATTAAGAGACCTTTGTGGATGATGCTTC 1067
781 ACTGTCCACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGAGAAATGAACACTG 840
1068 ACTGTCCACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGAGAAATGAACACTG 1127
841 CTCAATTGAGCCGACTGCTGGCTGGCACTGGCTGCAGTGTCTCAGCATTTTCCCAACA 900
1128 CTCATTGAGCCGACTGCTGGCTGGCACTGGCTGCAGTGTCTCAGCATTTTCCCAACA 1187
901 GTCTCTCCAGAACTAAGAACGCTCTCATTTGTACTCAGTGGGGGGAATGTAGACCTAAC 960
1188 GTCTCTCCAGAACTAAGAACGCTCTCATTTGTACTCAGTGGGGGGAATGTAGACCTAAC 1247
961 TCCTCTCAACTGGTGGGGGAGCTGAACGCCAGCTCTTACCAGAGCGGTCTGTTT 1016
1248 TCCCTGAACCTGGTGGGGAGCTGAACGCCAGCTCTTACCAGAGCGGTCTGTTT 1303
RESULT 3
AX224420
LOCUS AX224420 1020 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 3 from Patent WO0160987.
ACCESSION AX224420
VERSION AX224420.1 GI:15554662
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1020)
AUTHORS Meyers, R.A. and Rudolph-Owen, L.A.
TITLE A human pyridoxal-phosphate dependent enzyme family member and uses
therefor
JOURNAL Patent: WO 0160987-A 3 23-AUG-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
Location/Qualifiers
source 1..1020
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 288 a 220 c 243 g 269 t
ORIGIN
Query Match 79.4%; Score 807.8; DB 6; Length 1020;
Best Local Similarity 87.7%; Pred. No. 2.1e-211;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 1 ATGTGTGCTCAGTACGATCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCAA 60
Db 1 ATGTGTGCTCAGTACGATCTCTTGTGTGATGTTGAAAAGCTCATATCAACATTCAA 60
QY 61 GACTCTATCCACCTACCCAGTGTAAAGCTCCATTTTGAATCAATAGCAGGCGC 120
Db 61 GACTCTATCCACCTACCCAGTGTAAAGCTCCATTTTGAATCAATAGCAGGCGC 120
QY 121 AATCTTTTCTTCAATATGTGAGCTCTTCCAGAAATCGGTCTTTTAAAGATTCAGGTGCC 180
Db 121 AATCTTTTCTTCAATATGTGAGCTCTTCCAGAAATCGGTCTTTTAAAGATTCAGGTGCC 180
QY 181 CTTAATGCCATCAGAGGCTTAATTTCTGTACACGCCAGAGAGAAAGCCGAGTGT 240
Db 181 CTTAATGCCATCAGAGGCTTAATTTCTGTACACGCCAGAGAGAAAGCTGTTGTT 240
QY 241 ACTCAGCAGCGGGAACCATGCGCAAGCTCTCACCTATGCTCTAACTGGAGGAAT 300
Db 241 ACTCAGCAGCGGGAACCATGCGCAAGCTCTCACCTATGCTCTAACTGGAGGAAT 300
QY 301 CCGTCTTACATTTGGTTCCTCCAAACAGCTCCCAACTGCAAGAACTGGCAATCCAGCC 360
Db 301 CCGTCTTATATTGTTGGTGGCCCAAGAGCTCTCAAGAACTGGCAATCCAGCC 360
```

```

QY 361 TATGGAGCATCGATAGTATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 420
Db 361 TACGGAGCGTCAATTGTATCTGTGAACCTAGTGTAGTGCAGTCCAGAGAAAATGTTGCAAAA 420
QY 421 AGAATTATGCAAGAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAGTGATA 480
Db 421 AGAGTTTACAGAAGAAAGAGAGGATCATGTTATCATCCCAACAGGAGCCTGCAGTGATA 480
QY 481 GCTGGACAAGCAAAATTCCTGGAGTCTCAACACAGTTCCTTGGTGTAGATGACCTG 540
Db 481 GCTGGACAAGGACAATTCCTTGGAGTCTCAACACAGTTCCTTGGTGTAGATGACCTG 540
QY 541 GTGGTACCAAGTAGGAGGAGGAGGAGTGTGCTGGAATAGCCATTAACAATTAAGGCCCTG 600
Db 541 GTGGTACCTCTAGGTGGAGGAGGAGTGTGCTGGAATAGCAATTAAGGAGCTCTG 600
QY 601 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 660
Db 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATGCAGATGACTGCTACCAAGTCT 660
QY 661 AAACCTGAAGGAGCAACTGACCCCAATTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db 661 AAACCTGAAGGAGCAACTGACCCCAATTCATCTCCAGAAACCATAGCAGATGCTGTC 720
QY 721 AAATCCAGCATTCGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGATGCTTC 780
Db 721 AAATCCAGCATTCGCTTGAACACCTGGCCTATTATCAGGACCTTGTGGATGATGCTTC 780
QY 781 ACTGTACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGAGGAGATGAACACTG 840
Db 781 ACTGTACAGAGATGAATTAAGTGTGAACCCAGCTGTGTGGGGAGGAGATGAACACTG 840
QY 841 CTCATTGAGCGGACTGCTGGCGTGGCACTGGCTGTCAGTGTCTCAGCATTTCCAAACA 900
Db 841 CTCATTGAACTACAGCTGGTGTGGAGTGGCTGTCTGCTGTCTCAACATTTCAACT 900
QY 901 GTCTCTCCAGAAAGTAAAGACGCTCTGATGTACTCAGTGGGGGAGTGTAGACCTAA-- 958
Db 901 GTTTCCTCCAGAAAGTAAAGACATTTGATGTCTCAGTGGTGAATGTAGACTTTAACC 960
QY 959 -CCTCCCTGAACCTGGCTGGGAGGCTGAACGCCAGCTCTTACCAGAGCGTCTGTT 1016
Db 961 TCCTCCATACTTGGGTGAAGCAGGCTGAAAGCCAGCTCTTATCAGTGTGTTCTGT 1019

RESULT 4
LOCUS AY034081
DEFINITION Homo sapiens serine racemase mRNA, complete cds.
ACCESSION AY034081
VERSION AY034081.1
KEYWORDS GI:21307620
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1023)
AUTHORS Xia.M., Liu.Y. and Connolly.T.M.
TITLE Molecular cloning and expression of human serine racemase from
human NT2N cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1023)
AUTHORS Xia.M., Liu.Y. and Connolly.T.M.
JOURNAL Direct Submission
TITILE Submitted (08-MAY-2001) Pharmacology, Merck & Co., Sumneytown Pike,
West Point, PA 19486, USA
FEATURES
source Location/Qualifiers
1..1023
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13"
CDS
1..1023

```

```

/codon_start=1
/product="serine racemase"
/protein_id="AAK58495.1"
/db_xref="GI:21307621"
/translation="MCAQYCSIFADYVEKAHINIRDSIHLPVLTSSILNLTGRTNLF
KGELEFQTSFKIRGALNAVRSIVCEPDESRENVAKRTVEGIMVHPNEDAVI
YIVPOTAPDCKKLAIOYASIVCEPDESRENVAKRTVEGIMVHPNEDAVI
AGGGTALVNLQVLDALVVPVGGGMLAGIATVKAALKPSKVYAAEPSNADDCY
QSKLKGKLPNLYPPETIADGVKSSIGLNTWPIRLVDIDIFFVTEDEIKCATQLWE
RMSKLLEPTAGVGAVALSOHFQTVSPFVNICIVLGGNVDLTSSITWVQKAERPAS
YOSVSV"
BASE COUNT 290 a 220 c 243 g 270 t
ORIGIN
Query Match 79.4%; Score 807.8; DB 9; Length 1023;
Best Local Similarity 87.7%; Pred. No. 2.1e-211;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 1 ATGTGTGCTCAGTACTGCATCTCTTTTGTGTGATGTTTGAAGAAAGCTCATATCAACATTC 60
Db 1 ATGTGTGCTCAGTATTGCATCTCTTTTGTGTGATGTTTGAAGAAAGCTCATATCAACATTC 60
QY 61 GACTCTATCCACTCACCCAGTGTCAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 120
Db 61 GATTCTATCACCTCACAGCAGTGTCAACAAGCTCCATTTTGAATCAAACTAACAGGCGC 120
QY 121 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAAGCTGGGTCTTTTAAGATTCAGGTGCC 180
Db 121 AATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAAGCTGGGTCTTTTAAGATTCAGGTGCC 180
QY 181 CTTAATGCCATCAGAGGCTTAATCTCTGACACCCAGAGAGAGAGCCCAAGGCGGTAGTT 240
Db 181 CTCAATGCCCTCAGAGCTTGGTCTCTGATGCTTTAGAAAGAGAGCGGAAAGCTGTGTT 240
QY 241 ACTCACAGCAGCGGAAACCATGGCCAAAGCTCTCACCTATGCTGCTTAACTGGAAGAAAT 300
Db 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGAAAT 300
QY 301 CTGTCTTACATTTGTGTTCCCAACAGCTGCCCACTGCAAGAACTGGCAATTCACAGCC 360
Db 301 CTGTCTTATTTGTGTTGCCCAAGCAGCTGCCAGACTGTAAANAATTTGCAATACAGCC 360
QY 361 TATGGAGCATCGATAGTATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 420
Db 361 TACGGAGCGTCAATTCTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 420
QY 421 AGAATTATGCAAGAACAGAGGATCTTGGTCCATCCCAACAGGAGCCTGCAGTGATA 480
Db 421 AGAGTTTACAGAAGAAAGAGAGGATCATGGTATACATCCCAACAGGAGCCTGCAGTGATA 480
QY 481 GCTGGACAAGCAAAATTCCTTGGAGTGTCTGAACAGGTTCCCTTGGTGTAGATGACCTG 540
Db 481 GCTGGACAAGGACAATTCCTTGGAGTGTCTGAACAGGTTCCCTTGGTGTAGATGACCTG 540
QY 541 GTGGTACCAAGTAGGAGGAGGAGGAGTGTGCTGGAATAGCCATTAACAATTAAGGCCCTG 600
Db 541 GTGGTACCTCTAGGTGGAGGAGGAGTGTGCTGGAATAGCAATTAAGGAGCTCTG 600
QY 601 AAACCTAGTGTGAAGTATACGCTGTGAGCCCTCGAATGCAGATGACTGCTACCAAGTCT 660
Db 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATGCAGATGACTGCTACCAAGTCT 660
QY 661 AAACCTGAAGGAGCAACTGACCCCAATTCATCTCCAGAAACCATAGCAGATGCTGTC 720
Db 661 AAACCTGAAGGAGCAACTGACCCCAATTCATCTCCAGAAACCATAGCAGATGCTGTC 720
QY 721 AAATCCAGCATTCGCTTGAATACCTGGCCTATTATTAAGAGACCTTGTGGATGATGCTTC 780
Db 721 AAATCCAGCATTCGCTTGAACACCTGGCCTATTATCAGGACCTTGTGGATGATGCTTC 780
QY 781 ACTGTACCAAGATGAATCAAGTATGCAACCCAGCTGTGTGGGGAGGAGATGAACACTG 840
Db 781 ACTGTACAGAGATGAATTAAGTGTGAACCCAGCTGTGTGGGGAGGAGATGAACACTG 840
QY 841 CTCATTGAGCGGACTGCTGGCGTGGCACTGGCTGTCAGTGTCTCAGCATTTCCAAACA 900
Db 841 CTCATTGAACTACAGCTGGTGTGGAGTGGCTGTCTGCTGTCTCAACATTTCAACT 900
QY 901 GTCTCTCCAGAAAGTAAAGACGCTCTGATGTACTCAGTGGGGGAGTGTAGACCTAA-- 958
Db 901 GTTTCCTCCAGAAAGTAAAGACATTTGATGTCTCAGTGGTGAATGTAGACTTTAACC 960
QY 959 -CCTCCCTGAACCTGGCTGGGAGGCTGAACGCCAGCTCTTACCAGAGCGTCTGTT 1016
Db 961 TCCTCCATACTTGGGTGAAGCAGGCTGAAAGCCAGCTCTTATCAGTGTGTTCTGT 1019

```



TITLE A human pyridoxal-phosphate dependent enzyme family member and uses therefor

JOURNAL Patent: WO 0160987-A 1 23-AUG-2001;  
Millennium Pharmaceuticals, Inc. (US)

FEATURES  
source Location/Qualifiers  
1..1770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
69..1091  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC69570.1"  
/db\_xref="GI:15554661"  
/translation="MCAQYCSFADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLF  
KCELFQKTSFKIRGALNVRSLVDALPKPAVTHSGNHGQALTYAAKLEGPIA  
YIYVQTPADPKKLATQAGSIVCEPSEDSRENKVRTEGEGIMVHPNPEPAVI  
AGGQTALVELNVLPDVALVPGVGGMLAGIATIKALKPSKYVIAEAPSNAADQY  
OSKLGKLPNLYPPIETADGVKSSIGLNTWPIIRDLDVDFITVDEIKCATQLVWE  
RMKLLIEPTAGVGAVALSHQFQTVSPVKNICIVLSSGNNVLTSSITWVQKAERDAS  
YQSVSV"

BASE COUNT 499 a 379 c 410 g 482 t  
ORIGIN

Query Match 79.4%; Score 807.8; DB 6; Length 1770;

Best Local Similarity 87.7%; Pred. No. 2e-211;

Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 1 ATGTGTCAGTACGTCATCTCTTGGTGATGTTGAAAAAGCTATATCAACATCAA 60  
Db 69 ATGTGTCAGTATTCATCTCTTGGTGATGTTGAAAAAGCTATATCAACATTCGA 128  
QY 61 GACTCTATCCACCTCACCCAGTCTAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 120  
Db 129 GATTCTATCCACCTCACACCAGTGCTAACAGCTCCATTTTGAATCAAAATAGCAGGCGC 198  
QY 121 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACCTGGGTCTTTTAAAGTTCAGAGTGCC 180  
Db 189 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACCTGGGTCTTTTAAAGTTCAGTGCT 248  
QY 181 CTTAATGCCATCAGAGCTTAATCTCTGCACCCAGAGAGAGCCCAAGCCGTAGTT 240  
Db 249 CTCATGCGCTCAGAGCTTGGTCTCTGATGCTTTTAAAGAGGAGCGAAGCTGTGTT 308  
QY 241 ACTCACAGCAGCGAAACCACTGCGCAAGCTCTCACCTATGCTCTAACTGGAAGGAAT 300  
Db 309 ACTCACAGCAGTGAACCACTGCGCAGGCTCTCACTATGCTCCCAATTTGGAGGAAT 368  
QY 301 CTGCTTACATTTGGTTTCCCAAGACAGCTCCCACTGCAAGAACTGGCAATCCAGCC 360  
Db 369 CTTGCTTATATTGTTGGTGGCCAGACAGCTCCAGACTGTAAAAAACTTGAATACAGCC 428  
QY 361 TATGGAGCATCGATAGTATATCTGACCCCAAGTGACAGCTCCAGAGAAAGTCACTCA 420  
Db 429 TAGCGAGGTCGAATTTGATCTGAACTAGTGTGAGTGAGTGAGGAGGAAATGTTGCAAA 488  
QY 421 AGAATTATCAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGGCTGAGTGATA 480  
Db 489 AGAGTTACAGAAACAGAGAGGCATCATGGTACATCCCAACAGGAGGCTGAGTGATA 548  
QY 481 GCTGGACAGGACAATTTGCCCTGGAAGTCTGCAACCAAGTTCCCTTGGTAGTCACTG 540  
Db 549 GCTGGACAGGAGCAATTTGCCCTGGAAGTCTGCAACCAAGTTCCCTTGGTAGTCACTG 608  
QY 541 GTGGTACCAGTAGGAGGAGGAATGGTTGCTGGAATAGCCATTAACAATTAGGCCCTG 600  
Db 609 GTGGTACCAGTAGGAGGAGGAATGGTTGCTGGAATAGCAATTAACAATTAGGCCCTG 668  
QY 601 AAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCTACCACTCT 660  
Db 669 AAACCTAGTGTGAAGGTATATGCTGCTGAAACCTCAATGCAGATGACTGCTACCACTCT 728  
QY 661 AAACCTGAAGGAACTGACCCCAATCTTCATCTCCAGAAACCAATGACAGATGCTG 720

Db 729 AAGCTGAAGGGGAAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGT 788  
QY 721 AATCCAGCATTCGCTTGAATACCTGCCTTATTAAGAGACACCTTGTGGATGATCTTTC 780  
Db 789 AATCCAGCATTCGCTTGAACACCTGCCTTATTAAGAGACACCTTGTGGATGATATCTTC 848  
QY 781 ACTGTCCACCAAGATGAAATCAAGATGCAACCCAGCTGGTGGGGGAGAAATGAAACTG 840  
Db 849 ACTGTCCACAGAGATGAATTAAGTGTGCAACCCAGCTGGTGGGGAGAGATGAACACTA 908  
QY 841 CTCATTGAGCGGACTCTGCGGTGGCACTGGCTGCACTGCTGCTCAGCATTTCCAAACA 900  
Db 909 CTCATTGAACTACAGCTGGTGTGAGTGGCTGCTGCTGCTCCTCAACATTTTCAACT 968  
QY 901 GTCTCTCCAGAGTAAAGACGCTGCTCATTTGACTTACTCAGTGGGGGAGATGTAGACCTAA 958  
Db 969 GTTTCCTCCAGAGTAAAGAACATTTGTTGCTCAGTGGTGGTGAATGTAGACTTTAAC 1028  
QY 959 -CCTCCCTGAACCTGGGTGGCGAGCTGAACGCCAGCTCCTTTACCAGACGGTCTGTTT 1016  
Db 1029 TCCTCCATACCTGGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTCTGTTCTGT 1087

RESULT 7  
AK023169  
LOCUS Homo sapiens cDNA FLJ13107 fis, clone NT2RP3002501, weakly similar to THREONINE DEHYDRATASE CATABOLIC (EC 4.2.1.16).  
DEFINITION AK023169  
ACCESSION AK023169  
VERSION AK023169.1 GI:10434971  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens teratocarcinoma cell line: NT2 cDNA to mRNA, clone lib: NT2RP3 clone: NT2RP3002501.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Isozaki, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2477)  
Isodai, T. and Otsuki, T.  
Direct Submission  
Submitted (23-AUG-2000) Takao Isozaki, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert construction; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
FEATURES  
source Location/Qualifiers  
1..2477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002501"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP3"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."  
69..1091  
/note="unnamed protein product"  
/codon\_start=1  
CDS



/protein\_id="BAB14442.1"  
/db\_xref="GI:10434972"  
/translation="MCAOYCISFADVERAHINIRDSIHLTPVLTSILNLTGRLNLF  
KCELPQKTSFKIRGALANVSLVDALERPKAVVTHSSGNHGQALVYAAKLEGI  
YLVVPTAPDCKKLAIGASIVYCEPDESRENVAKRVYETEGIVHNPQERAVI  
ACQGHIALVNLQVLDALVYVGGGMLAGIAITVRAKLPSPVYVYAAEPPNADDCY  
QSKLKLMPNLPYPTETADGVKTSIGLNTWPIIRDLDVDFVTVEDEIKCATOLWE  
RKLKLIETAGVGAVALVLSQHFQTSVPEVKNICIVLSGNNVDLTSSITWVQARPA  
YQSVSV"  
BASE COUNT 738 a 535 c 524 g 680 t  
ORIGIN

Query Match 79.4%; Score 807.8; DB 9; Length 2477;  
Best Local Similarity 87.7%; Pred. No. 1.9e-211;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
QY 1 ATGTGTCACGACTGCTCCCTTGTGATGTGAAAGAGCTCATATCAACATTCAA 60  
DB 69 ATGTGTCACGATTTGCTATTCCTTGTGATGTGAAAGAGCTCATATCAACATTCGA 128  
QY 61 GACTCTATCCACCTCACCCAGTGTGTAACAAGCTCATTTTGAATCAATAGCAGGCGC 120  
DB 129 GATTCTATCCACTCACACGAGTGTGTAACAAGCTCATTTTGAATCAATAGCAGGCGC 188  
QY 121 AATCTTTTCTCAATGTGAGCTCTTCCAGAAACTGGGCTCTTTTAAAGATTGAGGTGCC 180  
DB 189 AATCTTTTCTCAATGTGAGCTCTTCCAGAAACTGGGCTCTTTTAAAGATTGAGGTGCC 248  
QY 181 CTTAATGCCATCAGAGGCTTATTTCTGACAGCCGAGAGAGAGAGCCCAAGCCGCTAGTT 240  
DB 249 CTTAATGCCGTCAGAGGCTTATTTCTGATGCTTTAGAAAGAGAGCCGAGAGCTGTGTT 308  
QY 241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTGTCTTAACTGCAAGCAATT 300  
DB 309 ACTCAGCAGGTGGAACCATGCGCAGGCTCTCACCTATGCTGTGCGCAATTTGGAAGGATT 368  
QY 301 CTGCTTACATTTGGTGTGCTCCCAACAGCTCCCAACTGCAAGAACTGCAATTCAGGCTG 360  
DB 369 CTGCTTATTTGTTGTCGCCAGCAGAGCTCCAGAGCTGTAAACAACTTGCATATCAAGCC 428  
QY 361 TATGAGCATCATAGTATGCTGACCAAGTGTGCTGCAAGTGTGAGAGAGAGGCTCACTCA 420  
DB 429 TACGAGCGCTCAATTTGATCTGTGAACCTAGTGTGATGAGTCCAGAGAAATTTGCAAAA 488  
QY 421 AGAATTTATCAAGAAACAGAGGCTCTTTGTGCTCCATCCCAACAGGAGGCTGCTGAGTGA 480  
DB 489 AGAGTTACAGAGAAACAGAGGCTCATGCTGATACATCCCAACAGGAGGCTGCTGAGTGA 548  
QY 481 GCTGACAGAGGAACATTTGCCCTGGAGTGTGACAGGTTCCCTTGGTAGATGACATG 540  
DB 549 GCTGACAGAGGAGCAATTTGCCCTGGAGTGTGCTGAACAGGTTCTCTTGGTGGATGCACTG 608  
QY 541 GTGGTACCAAGTAGGAGGAGGAGGATGTTGCTGGAATAGCCATACAAATTAAGGCCCTG 600  
DB 609 GTGGTACCTGTAGTGGAGGAGGAGGATGTTGCTGGAATAGCAATTTACAGTTAAGGCTG 668  
QY 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATGCAAGTCACTGCTACCAAGTCT 660  
DB 669 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATGCAAGTCACTGCTACCAAGTCT 728  
QY 661 AAACCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
DB 729 AAGCTGAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 788  
QY 721 AAATCCAGCATGGCTTGAATACCTGGCTTATTAAGAGACCTTTGTGGATCATCTCTC 780  
DB 789 AAATCCAGCATGGCTTGAATACCTGGCTTATTAAGAGACCTTTGTGGATCATCTCTC 848  
QY 781 ACTGTCACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
DB 849 ACTGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 908  
QY 841 CTCATTGAGCCGACTGTGCGGCTGGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

DB 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTGTCTCAACATTTTCAAACT 968  
QY 901 GTCTCTCCAGAGTAAAGAACGTCGCTATGTACTAGTGGGGGAGTATAGACCTAA-- 958  
DB 969 GTTTCCTCCAGAGTAAAGAACATTTGTATGTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
QY 959 -CTCTCCCTGAACCTGGGTGGGAGGCTGAACGCCAGCTCCCTTACCAGACGGTCTCTTT 1016  
DB 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAGGCCAGCTTCTTATCAGTCTGTTCTGT 1087  
RESULT 8  
LOCUS AX259575 1336 bp DNA linear PAT 26-OCT-2001  
DEFINITION Sequence 1 from Patent WO0173077.  
ACCESSION AX259575  
VERSION AX259575.1 GI:16508702  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE 1  
AUTHORS Ramakrishnan, S.  
TITLE Regulation of human serine racemase enzyme  
JOURNAL Patent: WO 0173077-A 1 04-OCT-2001;  
Bayer Aktiengesellschaft (DE)  
FEATURES  
source  
1. .1336  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 386 a 268 c 319 g 362 t 1 others  
ORIGIN  
Query Match 79.2%; Score 806.2; DB 6; Length 1336;  
Best Local Similarity 87.6%; Pred. No. 5.7e-211;  
Matches 893; Conservative 0; Mismatches 123; Indels 3; Gaps 1;  
QY 1 ATGTGTCCTCAGTCTGCTGCTCTCTTGTGATGTGAAAGAGCTCATATCAACATTCAA 60  
DB 52 ATGTGTCCTCAGTCTGCTGCTCTCTTGTGATGTGAAAGAGCTCATATCAACATTCGA 111  
QY 61 GACTCTATCCACCTCACCCAGTGTGTAACAAGCTCCATTTTGAATCAATAGCAGGCGC 120  
DB 112 GATTCTATCCACTCACACGAGTGTGTAACAAGCTCCATTTTGAATCAACAGGCGC 171  
QY 121 AATCTTTTCTCAATGTGAGCTCTTCCAGAAACTGGGCTCTTTTAAAGATTGAGGTGCC 180  
DB 172 AATCTTTTCTCAATGTGAGCTCTTCCAGAAACTGGGCTCTTTTAAAGATTGAGGTGCC 231  
QY 181 CTTAATGCCATCAGAGGCTTAAATTCCTGACAGCCGAGAGAGAGCCCAAGCCGCTAGTT 240  
DB 232 CTCATGCGCTCAGAGGCTTGGTCTCTGATGCTTTTGAAGAGGAGCCGAAAGCTGTGTT 291  
QY 241 ACTCAGCAGCGGAAACCATGCGCAAGCTCTCACCTATGCTGTCTTAACTGGAAGGAAAT 300  
DB 292 ACTCAGCAGGTGGAACCATGCGCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAAT 351  
QY 301 CCGTCTTACATTTGGTTCCTCCCAACAGCTCCCACTGCAAGAACTGCAATTCAGGCC 360  
DB 352 CCGTCTTATTTGTTGGTCCCGAGAGCTCCAGAGCTGTAAAGAACTTCAATACAGCC 411  
QY 361 TATGAGCATTCGATAGTACTGTGACCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
DB 412 TACGAGGCTCAATTTGATCTGTGAACCTAGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAG 471  
QY 421 AGAATTTATCAAGAAACAGAGGCTATTTGGTCTCCATCCCAACAGGAGGCTGAGTGA 480  
DB 472 AGAGTTACAGAGAAACAGAGGCTATGTTGATACATCCCAACAGGAGGCTGAGTGA 531  
QY 481 GCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540

```
Db 532 GTGGCAAGGAGCAATGTCCTGGAAGTCTGTAACAGGCTCTTTGGTGGATGCAGT 591
Qy 541 GTGTACAGCTAGGAGGAGGAGGAATGTTGCTGGAAATAGCCATTACAATTAAGGCCCTG 600
Db 592 GTGGTACCTGTAGTGGAGGAGGAGATGCTTGTGGAATACCAATACAGTTACAGGCTCTG 651
Qy 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATCAGATCAGACTCTACCAAGTCT 660
Db 652 AAACCTAGTGTGAAGGTATATGCTGTGAGCCCTCAATCAGATCAGACTCTACCAAGTCT 711
Qy 661 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTG 720
Db 712 AAGCTGAAGGGGAAACCTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTG 771
Qy 721 AAATCCAGCATGTGCTTGAATACCTGCGCTTATTATAGAGACCTGTGGATGATGCTTTC 780
Db 772 AAATCCAGCATGTGCTTGAACACCTGGCCCTATTATCAGGGACCTGTGGATGATGATCTTC 831
Qy 781 ACTGTCCACCAAGATGAATCAAGTATGCAACCCAGCTGGTGGGGGAGAAATGAACACTG 840
Db 832 ACTGTCCACAGAGTGAATTAAGTGTGCAACCCAGCTGGTGGGGAGAGATGAACACTA 891
Qy 841 CTATTCAGCCGACTGCTGGCGTGGCACTGGCTGAGTGTGCTCAGCATTTCCAAACA 900
Db 892 CTATTCAGCACTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 951
Qy 901 GTCTCTCCAGAAGTAAAGACGCTGCTATTGTACTCAGTGGGGGGAATGTAGACCTAA-- 958
Db 952 GTTTCCTCCAGAAGTAAAGACATTTGATTGTCTCAGTGGTGAATGTAGACTTTAACC 1011
Qy 959 -CTCCCTCAACTGGGTGGGCGAGCTGAAACGCCAGCTCCCTACCAGCGGTCTGTTT 1016
Db 1012 TCCTCCCACTAATTTGGGTGAAGCAGCGCTGAAAGGCCAGCTTCTTATCAGTCTGTTTCTGT 1070

RESULT 9
HSM805450
LOCUS Homo sapiens mRNA; cDNA DKFp762A2415 (from clone DKFp762A2415).
DEFINITION AL834378
ACCESSION AL834378.1 GI:21740048
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2264)
AUTHORS Ansoorge,W., Wirkner,U., Mewes,H.W., Weill,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
COMMENT Reseach Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFp762A2415) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
Location/Qualifiers
1..2264
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="22.9 cR from top of Chr17 linkage group"
/clone="DKFp762A2415"
/tissue_type="melanoma (Mewo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2264
/gene="DKFp762A2415"
<1..867
```

```
/gene="DKFp762A2415"
/notes="similarity to serine racemase (Mus musculus)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD39041.1"
/db_xref="GI:21740049"
/translation="GSFKIRGALNVAIRSLVLPDALKERKAVVTHSSGNHQAALYAAK
LEGIPAYIVVPTAPDCKKLAIOAGSIVYCPSPDESRENKRVYETEEFEGIMVHPN
QBPVATAGOGTALVNLQVPLVDALVVPVGGGMLAGTAITVKALKPSVKVYAAPPS
NADDCYQSKLKGKLPNLIPEPTIAGVYKSSIGLENTPWIRLDLDDIFVTVEDEIKCA
TOLVWERMKLLIEPTAGVGVAAVLSQHFOTVSPVKNKICIVLGGNVDLTSSITWVKQ
AERPASYQSVSV
2226..2231
/gene="DKFp762A2415"
2253
polyA_signal
polyA_site
BASE COUNT 692 a 478 c 465 g 629 t
ORIGIN
Query Match 65.3%; Score 664.6; DB 9; Length 2364;
Best Local Similarity 86.4%; Pred. No. 5.5e-172;
Matches 746; Conservative 0; Mismatches 114; Indels 3; Gaps 1;
Qy 157 GGCTCTTTAAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATTCCTGACACGCCA 216
Db 1 GGATCTTTTAAAGATTCGTGGTCTCTCAATGCGCTCAGAAGCTTGGTTCCTGATGCTTTA 60
Qy 217 GAAGAGAAGCCCAAGCCGTAGTTACTCAGCAGCGGGAACCATGCCAGCTCTCACC 276
Db 61 GAAAGGAAGCCGAAGAGCTGTTGTACTCAGCAGGTGGAACCATGCCAGGCTCTCACC 120
Qy 277 TATGCTGCTAACTGGAAGGAATTCCTGCTTACATTTGTTGTTCCCAAAACAGCTCCCAAC 336
Db 121 TATGCTGCCAATTTGGAAGGAATTCCTGCTTATTTGTTGTCGCCCAGACAGCTCCAGAC 180
Qy 337 TGAAGAAATCGCAATCCAAGCTATGGAGCATCGATAGTATATCTGTGACCCCAAGTGAC 396
Db 181 TGTAAAAAATCTGCAATACAGGCTACGGAGCGCTCAATTCGTATCTGTGAACCTAGTGT 240
Qy 397 GAGTCCAGAGAAAGTCACTCAAGAAATTTATCCAGAAACAGAGGATCTTTGGTCCAT 456
Db 241 GAGTCCAGAGAAATTTGCAAAAAGAGTTACAGAAGAAACAGAGGCAATCATGGTACAT 300
Qy 457 CCCAACAGGAGGCTGCAGTATAGCTGGACAAGAACAAATTCCTCGGAAGTGCCTGAAC 516
Db 301 CCCAACAGGAGGCTGCAGTATAGCTGGACAGGGACAAATTCGCCCTGGAGTGTCTGAAC 360
Qy 517 CAGGTTCCCTTGGTGTAGTGCACCTGGTGGTACCTGTAGTGGAGGAGGAATGGTTGTGGA 576
Db 361 CAGGTTCCCTTGGTGTAGTGCACCTGGTGGTACCTGTAGTGGAGGAGGAATGGTTGTGGA 420
Qy 577 ATAGCATTACAAATTAAGCCCTGAAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCG 636
Db 421 ATAGCAATTTACAGTTAAGGCTTGAACCTAGTGTGAAGGTATATGCTGTGCAACCTCA 480
Qy 637 AATGCAGATGACTGCTTACAGTCTAAACTGAAAGAGAGAACTGACCCCAATCTTCATCT 696
Db 481 AATGCAGATGACTGCTTACAGTCTCAAGTGAAGGGGAAACTGATGCCCAATCTTTATCT 540
Qy 697 CCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATA 756
Db 541 CCAGAAACCATAGCAGATGGTGTCAATCCAGCATTTGGCTTGAATACCTGGCCCTATTATC 600
Qy 757 AGAGACCTTGTGGATGATGCTTTCATCTGTCACCGAAGATCAATCAAGTATGCAACCCAG 816
Db 601 AGGACCTTGTGGATGATGATCTTTCATCTGTCACAGAGGATGAAATTAAGTGTGCAACCCAG 660
Qy 817 CTGGTGTGGGGAGAAATGAACTGCTCATTGACCGGACCTGCTGGCTGGGACCTGGCTGCA 876
Db 661 CTGGTGTGGGAGAGGATGAACTACTCATTGAACCTACAGCTGCTGTGGAGTGGCTGCT 720
Qy 877 GTGCTCTCAGCATTTCCAAACAGACTCTCTCCAGAAAGTAAAGAACCTGTGCAATGTACTC 936
|||||
```

```

Db      721 GTGCTGCTCAACATTTTCAAACTGTTTCCCGAGAGTAAAGAACATTTGTATTGTGCTC 780
QY      937 AGTGGGGGAATGTAGACATAA---CCTCCCTGAACCTGGGTGGGCGAGGCTGAAGGGCCA 993
Db      781 AGTGGTGAATGTAGACTTACCTCTCCATAACTTGGGTGAAGCAGGCTGAAGGGCCA 840
QY      994 GTCCTCTACACAGCGGTCTGTTT 1016
Db      841 GCTCTTATCAGTCTGTTTCTGT 863

RESULT 10
AL604066/c
LOCUS   AL604066        184151 bp    DNA        linear    ROD 14-MAR-2002
DEFINITION Mouse DNA sequence from clone Rp23-174M12 on chromosome 11,
complete sequence.
ACCESSION AL604066
VERSION   AL604066.9  GI:19571943
KEYWORDS  HTG.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 184151)
AUTHORS   Bray-Allen,S.
TITLE     Direct Submission
JOURNAL   Submitted (14-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Mar 21, 2002 this sequence version replaced gi:17902958.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Rp23-174M12 is
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone Rp23-174M12. The true
right end of clone Rp23-194P5 is at 82810 in this sequence.

FEATURES             source
     source
     1..184151
         /organism="Mus musculus"
         /db_xref="taxon:10090"
         /chromosome="11"
         /clone="Rp23-174M12"
         /clone_lib="RPI-23"
BASE COUNT  46879 a 41019 c 43482 g 52771 t
ORIGIN

Query Match      25.3%; Score 257.2; DB 10; Length 184151;
Best Local Similarity 73.1%; Pred. No. 6.1e-60;
Matches 421; Conservative 0; Mismatches 3; Indels 152; Gaps 1;

QY      593 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652
Db      125770 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 125711
QY      653 ACCAGTCTAACTGAAGAGAGAACTGACCCCAATCTTCATCCTCCAGAAACCATAGCAG 712

```

```

Db      125710 ACCAGTCTAACTGAAGAGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 125651
QY      713 ATGGTGTCAAAATCCAGCATTGCTTCAATACCTGGCCCTATTATAAGAGACCTTGTGGATG 772
Db      125650 ATGGTGTCAAAATCCAGCATTGCTTGAATACCTGGCCCTATTATAAGAGACCTTGTGGATG 125591
QY      773 ATGCTCTTCACTGTCAACCGAAGATGAATCA----- 802
Db      125590 ATGCTCTTCACTGTCAACCGAAGATGAATCAAGGTAAGACTCAGGCATTGTTAACAGAAC 125531
QY      803 ----- 802

Db      125530 CTGACAGTAGATCACATACTTATCACAAACCCAAATGTCTACTTCACTTTTTCGAATG 125471
QY      803 ----- 802

Db      125470 TAACTACTAGACTCCGGCTCCGTTCTTGTGTTTCAATTAACCCCTTGTCTCTTGTGGCCA 125411
QY      803 --ATGATGCAACCCAGCTGGTGTGGGGAGAAATGAAACTGCCTATTGAGCCGACTGCTGG 860
Db      125410 ACAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAAACTGCTCATTGACCCGACTGCTGG 125351
QY      861 CTGGCAGCTGGTGCAGTCTGTCTAGCATTTCCAAACAGCTCTCTCCAGAAAGTAAAGAA 920
Db      125350 CGTGGCAGCTGGTGCAGTCTGTCTAGCATTTCCAAACAGCTCTCTCCAGAAAGTAAAGAA 125291
QY      921 CGTGTGCATTGTACTCAGTGGGGGGAATGTAGACCTAACTCCCTGAACCTGGTGGGGA 980
Db      125290 CTTCTGCACTGTGACAGTGGGGGGAATGTAGACCTAACTCCCTGAACCTGGTGGGGA 125231
QY      981 GGCTGAACGGCCAGCTCTTACACACAGCGTCTGTTT 1016
Db      125230 GGCTGAACGGCCAGCTCTTACACACAGCGTCTGTTT 125195

RESULT 11
AC044847/c
LOCUS   AC044847        205910 bp    DNA        linear    HTG 07-DEC-2000
DEFINITION Mus musculus chromosome 11 clone Rp23-5305 map 11, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION AC044847
VERSION   AC044847.3  GI:11597106
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Mus musculus.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205910)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus chromosome 11, clone Rp23-5305
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 205910)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavski,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castelle,A., Choepe,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Piere,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,

```



misc_feature	/note="assembly_fragment" 21050..23372									
misc_feature	/note="assembly_fragment" 23473..26193									
misc_feature	/note="assembly_fragment" 26294..28324									
misc_feature	/note="assembly_fragment" 28425..31034									
misc_feature	/note="assembly_fragment" 31135..34444									
misc_feature	/note="assembly_fragment" 34545..38917									
misc_feature	/note="assembly_fragment" 39018..43628									
misc_feature	/note="assembly_fragment" 43729..47588									
misc_feature	/note="assembly_fragment" 47689..52019									
misc_feature	/note="assembly_fragment"									
Query Match 25.3%; Score 257.2; DB 2; Length 205910; Best Local Similarity 73.1%; Pred. No. 6e-60; Matches 421; Conservative 0; Mismatches 3; Indels 152; Gaps 1;										
QY	593	AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT	652							
Db	147884	AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT	147825							
QY	653	ACCAGTCTAACTGAAGGAGACTGACCCCAATCTTCATCTCCAGAACCATAGCAG	712							
Db	147824	ACCAGTCTAACTGAAGGAGACTGACCCCAATCTTCATCTCCAGAACCATAGCAG	147765							
QY	713	ATGGTGTCAAACTCAGCATTTGGCTTGAATACCTGGCCTATTATAAGAGACCTTTGGATG	772							
Db	147764	ATGGTGTCAAACTCAGCATTTGGCTTGAATACCTGGCCTATTATAAGAGACCTTTGGATG	147705							
QY	773	ATGCTCTTCACTGTCACCGAAGATGAATCA	802							
Db	147704	ATGCTCTTCACTGTCACCGAAGATGAATCAAGGTAAAGTAACTCAGGCATTGTTAACAGAACC	147645							
QY	803	-----	802							
Db	147644	CTGACAGTAGATCACATCTTATCACAAACCCAAATGTCTACTTCACTTTTTCGAATG	147585							
QY	803	-----	802							
Db	147584	TAACCTACTAGACTCCGGCTCCGTTTCTTTCATTAAACCTTGCTCTCTTCTGTGCCA	147525							
QY	803	--AGTATGCAACCCAGCTGGTGTGGGGGAGAAATGAACCTGCTCAATGAGCCGACTGCTGG	860							
Db	147524	ACAGTATGCAACCCAGCTGGTGTGGGGGAGAAATGAACCTGCTCAATGAGCCGACTGCTGG	147465							
QY	861	CGTGGCAGTGGCTGCTGCTCTCAGCATTTCCAAACAGTCTCTCCAGAGTAAGAA	920							
Db	147464	CGTGGCAGTGGCTGCTGCTCTCAGCATTTCCAAACAGTCTCTCCAGAGTAAGAA	147405							
QY	921	CGTCTGCATTTACTCAGTGGGGGGAATGTAGACCTTAACCTCCCTGAACCTGGTGGGCA	980							
Db	147404	CGTCTGCATTTACTCAGTGGGGGGAATGTAGACCTTAACCTCCCTGAACCTGGTGGGCA	147345							
QY	981	GGCTGAACGCCAGCTCTCTTACCAGACGGTCTGTTT	1016							
Db	147344	GGCTGAACGCCAGCTCTCTTACCAGACGGTCTGTTT	147309							
RESULT 12										
BC031531/c										
LOCUS										
DEFINITION										
Mus musculus. Similar to hypothetical protein FLJ10534, clone										
MGC:28320 IMAGE:4014819, mRNA, complete cds.										
ACCESSION										
BC031531										
VERSION										
BC031531.1 GI:21619397										
KEYWORDS										
MGC.										

SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 3379)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabpsr@mail.nih.gov">cgabpsr@mail.nih.gov</a> Tissue procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: <a href="mailto:ang@bcm.tmc.edu">ang@bcm.tmc.edu</a> Gunaratne, P. H., Garcia, A. M., Lu, X., Hulyk, S. W., Hale, S. M., Yoon, V. S., Kowis, C. R., Lawrence, S., Martin, R. G., Muzny, D. M., Richards, S., Gibbs, R. A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 36 Row: m Column: 16 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.	
FEATURES	Location/Qualifiers
1..3379	
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/map="C2ECH II"
	/clone="MGC:28320 IMAGE:4014819"
	/tissue_type="Mammary tumor metastasized to lung. MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."
	/clone_lib="NCI_CGAP_Lu30"
	/lab_host="DH10B"
	/note="Vector: pCMV-SPORT6"
	457..2187
CDS	/codon_start=1 /product="Similar to hypothetical protein FLJ10534" /protein_id="AAH31531.1" /db_xref="GI:21619398" /translation="MAKVADTILFLDPLEGWDSTGDYCLSCFLAOGLPYTLAVOGL SGFPKKQIDARKKLSKTVKRPEDKLLLDTOQESGMLRLANOKORHLAFRRR AVLFAHADVFPSEESDLVGTLLKISGYVRGRTLNVLNLSLHIVGHGFQNMQIDADVPDP FLNPRVTKSQKPNMAVEVCVTDAAADMEEDLVLMKADPDHQSLSQTEAIPDPMEG EQTWTEELDADLLKRSRVKVPKTSYQAEWILDEGSDSGEGGEYDIDIOH EGFMEESQDGGSEEEETMTLGESVRDLDLYDEKVEDEERMLEKYKQERLEEM FPDEMTPRDVAARIFQRYGLKSPRTSPWDPKENVLPNDYARIFQFNVNKRRI KTEKEAEGAEVGVVTLHVSQVSVVEYFQAGPLIAFSLPLPYEOKMSVLNVVS RNPGETPVPAKEELIFHCGRFRFRASPLFESQHTAADKKHFORFELTADALVTVFAP ITFPASVLLFQRRNGMSLIATGHLSFVDPDRWIKRVVLSGHPFKIFTKMAVVRY MFNFRGCDVV"
BASE COUNT	958 a 694 c 887 g 840 t
ORIGIN	
Query Match 24.6%; Score 250.8; DB 10; Length 3379; Best Local Similarity 72.4%; Pred. No. 5.3e-58; Matches 417; Conservative 0; Mismatches 7; Indels 152; Gaps 1;	
QY	593 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 652
Db	3223 AGGCCCTGAACCTAGTGTGAAGGTATACGCTGCTGAGCCCTCGAATGCAGATGACTGCT 3164
QY	653 ACCAGTCTAACTGAAGGAGAACTGACCCCAATCTTCACTCTCCAGAAACCATGACG 712

Db 3163 ACCAGTCTAACTGAAGGAGAACTGACCCCAATCTTCTCCAGAAACCATGACG 3104

QY 713 ATGGTGTCATCAATCCAGCATGGCTGTAATACCTGGCTATTATAGAGACCTTGTGGATG 772

Db 3103 ATGGTGTCATCAATCCAGCATGGCTGTAATACCTGGCTATTATAGAGACCTTGTGGATG 3044

QY 773 ATGCTCTTCACTGTCACCGGAAGATGAATCA----- 802

Db 3043 ATGCTCTTCACTGTCACCGGAAGATGAATCAAGCTAGACTCAGGCATTGTTAACAGAAC 2984

QY 803 ----- 802

Db 2983 CTGACAGTAGATCACATACCTATTATACAAACCAACCAAAATGCTACCTCATTTTTCGAATG 2924

QY 803 ----- 802

Db 2923 TAACCTACTAGACTCGGCTCCGTTCTTGTTTCATTAACCCCTTGCTCTTCTGTGCCA 2864

QY 803 --AGTATGCAACCCAGCTGGTGTGGGGAGAAATGAACCTGCTCATTCAGCCGACTGCTGG 860

Db 2863 ACAGTGTGCAACCCAGCTGGTGTGGGGAGAAATGAAGCTGCTCATTCAGCCGACTGCTGG 2804

QY 861 CGTGGCACTGGCTGCTGCTCAGCATTTTCCAAACAGTCTCTCCAGAACTAAGAA 920

Db 2803 CGTGGCACTGGCTGCTGCTCAGCATTTTCCAAACAGTCTCTCCAGAACTAAGAA 2744

QY 921 CGTCTGCTACTGCTCAGTGGGGGAGATGAGACCTAACCTCCCTGAACCTGGGTGGGCA 980

Db 2743 CGTCTGCTACTGCTCAGTGGGGGAGATGAGACCTAACCTCCCTGAACCTGGGTGGGCA 2684

QY 981 GCGTGAACGCCAGCTCCTTACCAGACGGTCTCTGTT 1016

Db 2683 GCGTGAACGCCAGCTCCTTACCAGACGGTCTCTGTT 2648

## RESULT 13

AC122087/c  
LOCUS AC122087 101814 bp DNA linear HTG 23-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-280H23, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 61 unordered pieces.  
ACCESSION AC122087  
VERSION AC122087.2 GI:21903132  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 101814)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J.,  
Hernandez,O., Hodson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homs,I.F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameris,K., Tang,H.,  
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G., and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 101814)  
Worley,K.C.  
Direct Submission  
Submitted (22-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 101814)  
Worley,K.C.  
Direct Submission  
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 18, 2002 this sequence version replaced gi:21070489.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GW21  
Center clone name: CH230-280H23  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 43731 bases at least Q40  
Consensus quality: 47905 bases at least Q30  
Consensus quality: 51501 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1045: contig of 1045 bp in length  
\* 1046 1145: gap of unknown length  
\* 1146 2200: contig of 1055 bp in length  
\* 2201 2300: gap of unknown length  
\* 2301 3345: contig of 1045 bp in length  
\* 3346 3445: gap of unknown length  
\* 3446 4445: contig of 1000 bp in length  
\* 4446 4545: gap of unknown length  
\* 4545 5575: contig of 1030 bp in length  
\* 5575 5675: gap of unknown length  
\* 5676 6696: contig of 1021 bp in length  
\* 6697 7807: contig of 1011 bp in length  
\* 7808 7907: gap of unknown length  
\* 7908 8910: contig of 1003 bp in length  
\* 8911 9010: gap of unknown length

TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 101814)  
AUTHORS  
Worley,K.C.  
TITLE  
JOURNAL  
REFERENCE  
3 (bases 1 to 101814)  
AUTHORS  
Worley,K.C.  
TITLE  
JOURNAL  
COMMENT





Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3807)  
Strausberg, R.  
Direct Submission  
Submitted (07-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-re@mail.nih.gov](mailto:cgapsb-re@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nigri.nih.gov](mailto:nisc_mgc@nigri.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL plate: 39 Row: j Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922495.

## FEATURES

Source

Location/Qualifiers  
1..3807  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:29514 IMAGE:4896567"  
/tissue\_type="Brain, neuroblastoma"  
/clone\_lib="NIH\_MGC\_19"  
/lab\_host="PH10B-R"  
/note="Vector: pOTB7"  
319..2289  
/codon\_start=1  
/product="Unknown (protein for MGC:29514)"  
/protein\_id="AAH19090.1"  
/db\_xref="GI:17512218"  
/translation="MAKVADTILFLDPLEGWDSTGDCYLSCLFAGQLPTTYTLAVQGI  
SGLPKQIDTRKLSKAVEKRFPHDKLLDLDQEQEAGMLRLQLANQOHLAFDRR  
AYLFAHVDVFPSENNLVGTLKISGYRGOTLVNRLHIVYGDGFQMKQIDAPDGP  
FPLNPRGKPKQDPDAMEICATDAVDOMEGLKVLKADPGRQESLOAEVTPDPMEG  
EOTWTEELSEAKDFLEKGSVKVKKVPKTSYQAEWLDGSGSGGEGDEYEDDM  
EHEDMEESQDSESEEEYEYTWIGESVHDDLDYKVDDEAEAKMLEKYQERLEE  
MPDEVDTPDVAARIRQKYRGLKSFRTSPWDPKLENLPQDYARIFQFQNTNTNRKI  
FKEVEKEVEAGVYVTLHVSEVPSVVEFCRQGTPLIAFSLPHQKMSVLNNV  
RRDPGTEVKAKEELIFHCGFRFRFPASPLFSQHTAADKHLQRELTADMLAVIYA  
PTTFPASVLLKQKNSGMSHLIATGSLHMSVDPDRMWTKRVVLSGHPKFIETKMAVVR  
YMFENRDVLFKPKVELRTKWRGRHKEPLGLTHGHMKCSFDGKLKSQDTVLNLYKR  
VFPKTYDYPYEPVPWLKSEISSVPOGGME"

## CDS

1102 a 813 c 943 g 949 t

## BASE COUNT

ORIGIN

Query Match 17.9%; Score 181.8; DB 9; Length 3807;  
Best Local Similarity 88.0%; Pred. No. 5.3e-39;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 593 AGGCCCTGAACCTAGTGTGAAGGTATAGCTGTGAGCCCTCGAATGCAGATGACTGCT 652  
|||||  
Db 3351 AGGCTCTGAACCTAGTGTGAAGGTATAGCTGTGAGCCCTCGAATGCAGATGACTGCT 3292  
QY 653 ACCAGCTCTAACTGAAGAGGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
|||||  
Db 3291 ACCAGTCCAAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 3232

QY 713 ATGGTGTCAAATCCACGACATTGGCTTCAATACCTGGCTATTATTAAGAGACCTTGTGGATG 772  
|||||  
Db 3231 ATGGTGTCAAATCCACGACATTGGCTTCAACACCTGGCTATTATTCAGGGACCTTGTGGATG 3172  
QY 773 ATGCTCTTCACTGTCAACCGAAGATGAATCAAGTATGCAACCCAGC 817  
|||||  
Db 3171 ATATCTTCACTGTCAACAGAGATGAATTAAGGTGAGGCTCCAGC 3127  
RESULT 15  
AK026565/c  
LOCUS  
DEFINITION  
Homo sapiens cDNA: FLJ22912 fis, clone KAT06132. PRI 29-SEP-2000  
AK026565  
ACCESSION  
AK026565.1 GI:10439447  
VERSION  
o1igo capping; fis (full insert sequence).  
KEYWORDS  
Homo sapiens signet-ring cell carcinoma cell\_line:KATO III cDNA to  
SOURCE  
mRNA, clone\_lib:KAT clone:KAT06132.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (sites)  
AUTHORS  
Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,  
Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
TITLE  
Unpublished  
JOURNAL  
2 (bases 1 to 4140)  
REFERENCE  
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
TITLE  
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: [cdna@ims.u-tokyo.ac.jp](mailto:cdna@ims.u-tokyo.ac.jp), Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)  
COMMENT  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
Location/Qualifiers  
Source  
1..4140  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="KAT06132"  
/cell\_line="KATO III"  
/cell\_type="signet-ring cell carcinoma"  
/clone\_lib="KAT"  
/note="cloning vector pME18SFL3"  
BASE COUNT 1195 a 901 c 1050 g 994 t  
ORIGIN  
Query Match 17.9%; Score 181.8; DB 9; Length 4140;  
Best Local Similarity 88.0%; Pred. No. 5.3e-39;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 593 AGGCCCTGAACCTAGTGTGAAGGTATAGCTGTGAGCCCTCGAATGCAGATGACTGCT 652  
|||||  
Db 3507 AGGCTCTGAACCTAGTGTGAAGGTATAGCTGTGAGCCCTCAATGCAGATGACTGCT 3448  
QY 653 ACCAGCTCTAACTGAAGAGGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
|||||  
Db 3447 ACCAGTCCAAGCTGAAGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAG 3388  
QY 713 ATGGTGTCAAATCCACGACATTGGCTTGAATACCTGGCTTATTATTAAGAGACCTTGTGGATG 772  
|||||  
Db 3387 ATGGTGTCAAATCCACGACATTGGCTTGAACACCTGGCTTATTATTCAGGGACCTTGTGGATG 3328  
QY 773 ATGCTTCACTGTCAACCGAAGATGAATCAAGTATGCAACCCAGC 817



Db 3327 ATATCTCACTGTCACAGAGGATGAAATTAAGGTGAGGCTCCAGC 3283  
|||||  
|||||

Search completed: June 24, 2003, 05:10:53  
Job time : 2676 secs

**THIS PAGE BLANK (USPTO)**





XX The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
XX Sequence 1672 BP; 491 A; 394 C; 379 G; 408 T; 0 other;  
SQ

Query Match 99.3%; Score 1011.2; DB 21; Length 1672;  
Best Local Similarity 99.7%; Pred. No. 5.9e-312;  
Matches 1013; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTGTCAGTACTGATCTCTTCTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
DB 219 ATGTGTCAGTACTGATCTCTTCTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 278  
QY 61 GACTCTATCCACCTCACCCAGTGTCTAACAAGCTCATTTGAATCAATAGCAGGCGC 120  
DB 279 GACTCTATCCACCTCACCCAGTGTCTAACAAGCTCATTTGAATCAATAGCAGGCGC 338  
QY 121 AATCTTTTCTTCAAAATGTAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCGAGGTGCC 180  
DB 339 AATCTTTTCTTCAAAATGTAGCTCTTCCAGAAAACCTGGGTCTTTTAAAGATTCGAGGTGCC 398  
QY 181 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCAGAGAGAACCCCAAGCCGTAGTT 240  
DB 399 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCAGAGAGAACCCCAAGCCGTAGTT 458  
QY 241 ACTCACAGCGGGAACCATGGCCAAAGCTCTCACTATGCTGCTAAACTGGAAGGAAT 300  
DB 459 ACTCACAGCGGGAACCATGGCCAAAGCTCTCACTATGCTGCTAAACTGGAAGGAAT 518  
QY 301 CTGCTTACATGTTGTTTCCCAAAACAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 360  
DB 519 CTGCTTACATGTTGTTTCCCAAAACAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 578  
QY 361 TATGAGCATCGATAGTATGTCACCAAGTGCAGAGTCAGAGAAAGGTCACTCAA 420  
DB 579 TATGAGCATCGATAGTATGTCACCAAGTGCAGAGTCAGAGAAAGGTCACTCAA 638  
QY 421 AGAATTATGCAAGAAACAGAGGCAATCTTGCTCAATCCCAACAGAGGCTGCAGTGATA 480  
DB 639 AGAATTATGCAAGAAACAGAGGCAATCTTGCTCAATCCCAACAGAGGCTGCAGTGATA 698  
QY 481 GCTGACAAGGAACAATTCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATCACTG 540  
DB 699 GCTGACAAGGAACAATTCCTGGAAGTGTGTAACCAAGGTTCCCTTGGTAGATCACTG 758  
QY 541 GTGGTACAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATACAAATTAAGCCCTG 600  
DB 759 GTGGTACAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATACAAATTAAGCCCTG 818  
QY 601 AAACCTAGCTGCAAGGTATAGCTGCTGAGCCCTGGAATGCAATGAGTGCCTACCAAGTCT 660  
DB 819 AAACCTAGCTGCAAGGTATAGCTGCTGAGCCCTGGAATGCAATGAGTGCCTACCAAGTCT 878  
QY 661 AAATGAAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATACAGATGGTCTC 720  
DB 879 AAATGAAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATACAGATGGTCTC 938  
QY 721 AAATCCAGATGGCTTGAATACCTGGCTTATATAAGAGACCTTGTGGATGATGCTTC 780  
DB 939 AAATCCAGATGGCTTGAATACCTGGCTTATATAAGAGACCTTGTGGATGATGCTTC 998  
QY 781 ACTGTCACCGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGGAATGAACCTG 840  
DB 999 ACTGTCACCGAAGATGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGGAATGAACCTG 1058

QY 841 CTCATTGAGCCGACTGCTGGCTGGCACTGGCTGCTCAGTGTCTCTCAGCATTTCCAAACA 900  
DB 1059 CTCATTGAGCCGACTGCTGGCTGGCACTGGCTGCTCAGTGTCTCTCAGCATTTCCAAACA 1118  
QY 901 GCTCTCTCAGAAAGTAAAGAACCTCTGCAATGTACTCAGTGGGGGAATGTAGACCTAAC 960  
DB 1119 GCTCTCTCAGAAAGTAAAGAACCTCTGCAATGTACTCAGTGGGGGAATGTAGACCTAAC 1178  
QY 961 TCCCTGAACTGGGTGGGCGAGGCTCAACGGCCAGCTCTTACCACAGCGTCTGT 1016  
DB 1179 TCCCTGAACTGGGTGGGCGAGGCTCAACGGCCAGCTCTTACCACAGCGTCTGT 1234

RESULT 3  
AAA59299  
ID AAA59299 standard; DNA; 1023 BP.  
XX  
AC AAA59299;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE DNA encoding a human serine racemase polypeptide.  
XX  
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1023  
FT /\*tag= a  
FT /product= "serine racemase"  
FT  
XX  
PN WO200043526-A1.  
XX  
PD 27-JUL-2000.  
XX  
PF 18-JAN-2000; 2000WO-US00938.  
XX  
PR 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX  
PA (UJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
XX  
PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
PI Ferris CD;  
XX  
DR WPI; 2000-482915/42.  
DR P-PSDB; AAB07734.  
XX  
PT Mammalian serine racemase preparations, used to identify modulators  
PT which can be used to treat diseases associated with  
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
PT  
XX  
PS Claim 19; Page 49-50; 54pp; English.  
XX  
CC The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
SQ Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;

Query Match		79.4%; Score 807.8; DB 21; Length 1023;
Best Local Similarity		87.7%; Pred. No. 4.1e-247;
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;		
QY	1	ATGTGTGCTCAGTACTGTCATCTCCCTTTTGTGTGATGTTGAAAAAGCTCATATCAACATCAA 60
DB	1	ATGTGTGCTCAGTATTGTCATCTCCCTTTTGTGTGATGTTGAAAAAGCTCATATCAACATCGA 60
QY	61	GACTCTATCCACTCACCCTCAGTCTTAACAAGCTCCATTTGAATCAATACAGGGGCG 120
DB	61	GATCTATCCACTCACCCTCAGTCTTAACAAGCTCCATTTGAATCAATACAGGGGCG 120
QY	121	AATCTTTTCTTCAATATGTGAGTCTTCCAGAAACTGGCTCTTTAAAGATTCAGGTGCC 180
DB	121	AATCTTTTCTTCAATATGTGAATCTTCCAGAAACAGGATCTTTAAGATTCGTGGTCT 180
QY	181	CTTAATGTCATCAGAGGCTTAATCTCTGACAGCCGAGAGAAAGCCCAAGCCGTAGTT 240
DB	181	CTCAATGCCGTGAGAAGCTTGGTCTCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 240
QY	241	ACTCAGAGCGGGAACCATGGCCAGGCTCTCAGCTATGCTTAACTGGAAGGAATT 300
DB	241	ACTCAGAGCGGGAACCATGGCCAGGCTCTCAGCTATGCTTAACTGGAAGGAATT 300
QY	301	CCTGCTTACATGTGTGTTCCCAAAAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 360
DB	301	CCTGCTTATATTGTGTGCCCCAGACAGCTCCAGACTGTAAAAAACTTCCATACAAGCC 360
QY	361	TATGAGAGTCGATAGTACTGTGACCCAGTACGAGTCCAGAGAAAGGTCACTCAA 420
DB	361	TACGGAGGCTCAATTGTACTGTGAACCTAGTAGTATGATGAGTCCAGAGAAATGTTCAAAA 420
QY	421	AGAATTATGCAAGAACAGAGGCAATCTGTGTCATCCCAAGGAGGCTGCACTGATA 480
DB	421	AGAGTTACAGAGAACAGAGGCAATCTGTGTCATCCCAAGGAGGCTGCACTGATA 480
QY	481	GCTGACAGGAAACAATTGCCCTGGAAGTGTGTAACCAAGTTCCTTGTGTAGATCACTG 540
DB	481	GCTGACAGGAAACAATTGCCCTGGAAGTGTGTAACCAAGTTCCTTGTGTAGATCACTG 540
QY	541	GTGGTACCAAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTACAAATTAAGCCCTG 600
DB	541	GTGGTACCTGTAGGTGAGGAGGAGGAATGTTGCTGGAATAGCAATTAAGGCTCTG 600
QY	601	AAACCTAGTGTGAAGGTATACGCTCTGAGCCCTGCAATGCAAGTACGCTGCTACCACT 660
DB	601	AAACCTAGTGTGAAGGTATATGCTGCTGAACCTCAATGCAAGTACGCTGCTACCACT 660
QY	661	AAACTGAAAGGAGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTCT 720
DB	661	AAAGCTGAAAGGAGAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGGTCT 720
QY	721	AAATCCAGATTGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATCTCTTC 780
DB	721	AAATCCAGATTGGCTTGAATACCTGGCTTATTAAGAGACCTTGTGGATGATCTCTTC 780
QY	781	ACTGTCCAGGAGTGAATCAAGTATGCAACCCAGCTGGTGTGGGAGAGATGAACCTG 840
DB	781	ACTGTCCAGGAGTGAATTAAGTGTGCAACCCAGCTGGTGTGGGAGAGATGAACCTA 840
QY	841	CTCATTGAGCCGACTGCTGGGTGCACTGGCTGAGTCTGCTCAGCAATTTCCAAACA 900
DB	841	CTCATTGAACCTTACAGCTGGTGTGGAGTGGCTGCTGCTCTCAACATTTTCAAACT 900
QY	901	GTCCTCCAGAGTAAAGCAAGCTCTGATTTGATCTAGTGGGGGAAATGTAGACCTAA-- 958
DB	901	GTTCCTCCAGAGTAAAGCAATTTGATTTGCTGCTGAGTGGGAAATGTAGACTTAAC 960
QY	959	-CCTCCCTGAACTGGGTGGGAGGCTGAACGGCCAGCTCCTTACCAGAGGCTCTGTTT 1016
DB	961	TCCTCCATAACTTGGGTGAACGAGGCTGAAGGCCAGCTCTTATCATGCTGTTTCTGT 1019

RESULT 4	
AAS15217	
ID	AAS15217 standard; cDNA; 1023 BP.
XX	
AC	AAS15217;
DT	16-JAN-2002 (first entry)
XX	
DE	Human cDNA encoding Serine Racemase.
XX	
KW	Human; ss; serine racemase; Parkinson's disease; Huntington's disease; anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia; schizophrenia; transgenic animal; chromosome 17p13.
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	1..1023
FT	/*tag= a
FT	/product= "Serine racemase"
XX	
PN	WO200175144-A1.
XX	
PD	11-OCT-2001.
XX	
PF	02-APR-2001; 2001WO-US10662.
XX	
PR	04-APR-2000; 2000US-194451P.
XX	
PA	(MERI ) MERCK & CO INC.
XX	
PI	Connolly T, Liu Y, Xia M;
XX	
DR	WPI: 2001-656991/75.
DR	P-PSDB; AAU09124.
XX	
PT	New recombinant serine racemase polypeptide, useful in assays for identifying compounds that alter enzyme activity (e.g. including compounds that inhibit or stimulate enzyme activity) or in generating antibodies against the protein
PT	
PT	
PT	
XX	
PS	Claim 1; Page 9; 43pp; English.
XX	
CC	The invention relates to a recombinant polynucleotide encoding a human serine racemase, vectors containing it, host cells expressing the racemase, methods of identifying inhibitors of serine racemase and a transgenic animal lacking a functional endogenous serine racemase comprising the human serine racemase of the invention. The racemase protein is useful in assays to identify compounds that inhibit or stimulate enzyme activity, in the generation of antibodies against the protein, and in structural studies of the protein and
CC	structure/function relationships of the protein. Biologically active fragments, and mutant or polymorphic forms of the serine racemase polypeptide have diagnostic, therapeutic or prophylactic uses (e.g. for neurological diseases such as Parkinson's and Huntington's disease, anxiety, glaucoma, stroke, hyperalgesia, pain, spinocerebellar ataxia and schizophrenia), and would be useful for screening for modulators and/or inhibitors of serine racemase function. The polynucleotides are useful as probes for the specific detection of the presence of a polynucleotide encoding a serine racemase protein, and as primers for nucleic acid amplification based assays for the detection of polynucleotides encoding serine racemase protein. The transgenic animal is useful for the study of the tissue and temporal specific expression or activity of the serine racemase gene in an animal. The gene for human serine racemase is located on chromosome 17p13. The present sequence encodes human serine racemase.
CC	
CC	Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;
SQ	

Query Match 79.4%; Score 807.8; DB 22; Length 1023;  
Best Local Similarity 87.7%; Pred. No. 4.1e-247;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;



Db 129 GATTCTATCCACCTCACACAGTGTCTAACAGTCCATTTTGAATCAACTAACAGGGCGC 188  
Qy 121 AATCTTTTCTCAAAATGAGTCTTCCAGAAACTGGTCTTTTAAGATTCGAGGTGCC 180  
Db 189 AATCTTTTCTCAAAATGAGTCTTCCAGAAACAGGATCTTTTAAGATTCGAGGTGCC 248  
Qy 181 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCGAGAGAGAGCCCAAGCGGTAGTT 240  
Db 249 CTCATGCCCTCAGAGCTTGGTTCCTGTATGCTTTAGAAAGGAGCGGAAGCTGTGTT 308  
Qy 241 ACTCAGCAGCGGGAACCATGGCCAGCTCTCACCTATGCTGCTAACTGGAGGAATT 300  
Db 309 ACTCAGCAGCTGGAACCATGGCCAGCTCTCACCTATGCTGCTAACTGGAGGAATT 368  
Qy 301 CTTGCTTACATGTGTTTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCCAAGCC 360  
Db 369 CTTGCTTATATGTGTTGCCCCAGACAGCTCCAGACTGTAAACACTTGCATACAGCC 428  
Qy 361 TATGGAGCATCGATAGTATATCTGTGACCCCAAGTGACAGTCCAGAGAAAAGTCACTCAA 420  
Db 429 TACGGAGCGTCAATTTGTATCTGTGAACCTAGTGTGATGAGTCCAGAGAAAATGTTCAAAA 488  
Qy 421 AGAATATGCAAGAAACAGAGGATCTTGGTCCATCCCAACAGGAGCCCTGCAGTGATA 480  
Db 489 AGATTTACAGAAACAGAGGATCATGTGTATATCCCAACAGGAGCCCTGCAGTGATA 548  
Qy 481 GCTGGCAAGGAACAATTTGCCCTGGAGTGTCTGAACCCAGCTTCCCTGGTGTAGTCACTG 540  
Db 549 CTTGGACAGGGACAATTTGCCCTGGAGTGTCTGAACCCAGGTTCTTTGGTGGATGCACTG 608  
Qy 541 GTGTTACAGTAGGAGGAGGAATGTTGCTGGAATAGCCATTACAATTAAGGCCCTG 600  
Db 609 GTGTTACCTCTAGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTG 668  
Qy 601 AATCTAGTGTAGGTATAGTGTGAGCCCTCGAATCGAGATGACTGCTACCACTCT 660  
Db 669 AATCTAGTGTAGGTATATGCTGTGTAACCCCTCAATCGAGATGACTGCTACCACTCT 728  
Qy 661 AATCTAAGAGGAACATGACCCCAATCTTTCATCTCCAGAAACCATAGCAGATGTTGTC 720  
Db 729 AGCTGAAGGGGAACATGATGCCCATCTTTATCTCCAGAAACCATAGCAGATGTTGTC 788  
Qy 721 AATCCAGCATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTGTGGATGATCTTC 780  
Db 789 AATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTGTGGATGATCTTC 848  
Qy 781 ACTGTCAACCAAGATGAATCAAGTATGCAACCCAGCTGTTGTGGGGAGAAATGAACCTG 840  
Db 849 ACTGTCAACCAAGATGAATTAAGTGTGCAACCCAGCTGTTGTGGGGAGAAATGAACCTG 908  
Qy 841 CTCATTGAGCCGACTGCTGGCGTGGCACTGGCTGCAGTGTCTGCTCAGCATTTTCCAAACA 900  
Db 909 CTCATTGAACCTACAGCTGTTGTTGGAGTGGCTGCTGCTCTCAACATTTTCAACT 968  
Qy 901 GTCTCTCCAGAAATGAAGACGCTGCTGATTTGTTACTCAGTGGGGGAATGTAGACCTAA-- 958  
Db 969 GTTTTCCCAAGAAATAAGAACATTTGTTATGCTCAGTGGTGAATGTAGACTTTAACC 1028  
Qy 959 -CTCTCCCTGNACTGGGTGGGCGAGGCTGACAGCCGAGCTCTTACCAGACGGTCTGTT 1016  
Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAGCCAGCTCTTATCAGTCTGTTTCTGT 1087

RESULT 6  
AAH16282  
ID AAH16282 standard; cDNA; 2477 BP.  
XX  
AC AAH16282;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15148.

XX  
KW Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isodai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 15148; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2477 BP; 738 A; 535 C; 524 G; 680 T; 0 other;

Query Match 79.4%; Score 807.8; DB 22; Length 2477;  
Best Local Similarity 87.7%; Pred. No. 6.6e-247;  
Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;  
Qy 1 ATGTGTGCTCAGTACTGCATCTCTTTGCTGTGATGTTTGAAGAAAGCTCATATCAACATTCAA 60  
Db 69 ATGTGTGCTCAGTACTGCATCTCTTTGCTGTGATGTTTGAAGAAAGCTCATATCAACATTCGA 128  
Qy 61 GACTCTATCCACTCACCTCACCCAGTGCCTAACAGCTCCATTTTGAATCAATAGCAGGGCGC 120  
Db 129 GATTCTATCCACTCACCTCACCCAGTGCCTAACAGCTCCATTTTGAATCAACAGGGCGC 188  
Qy 121 AATCTTTTCTCAAAATGAGTCTTCCAGAAACTGGTCTTTTAAGATTCGAGGTGCC 180  
Db 189 AATCTTTTCTCAAAATGAGTCTTCCAGAAACAGGATCTTTTAAGATTCGAGGTGCC 248  
Qy 181 CTTAATGCCATCAGAGGCTTAATTCCTGACAGCCGAGAGAAAGCCGAGTGT 240



Db 249 CTCATGCGTCAGAGCTGGTCTCTGATGCTTTAGAAAGAGAGCCGAAAGCTGTTGTT 308  
 QY 241 ACTCACAGCAGCAAGCACTGGCCAAAGCTCTCACCTATGCTGCTAAACTGGAAGGAATT 300  
 Db 309 ACTCACAGCAGTGAACCACTGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGGAATT 368  
 QY 301 CCTGCTTACATGTGGTTCCTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCAGCC 360  
 Db 369 CCTGCTTATATGTGGTCCCGCAGACAGCTCCAGAGCTGTAAAGAACTTTGCAATACAGCC 428  
 QY 361 TATGAGCATCGATAGTATATGTCAGCCCAAGTGCAGAGTCAGAGAAAGAGTCACTCAA 420  
 Db 429 TACGAGCGCTCAATGTATGTAACCTAGTAGTATGATGAGTCCAGAGAAATTTGCAAAA 488  
 QY 421 AGAATATTCAGAAACAGAGGATCTTGTGTCATCCCAACAGGAGCCCTGCAGTGATA 480  
 Db 489 AGAGTTACAGAAAGAACAGAGGATCATGTACATCCCAAGGAGCCCTGCAGTGATA 548  
 QY 481 GCTGACAAAGAACAAATTCCTGGAAGTGTGTAACAGGTTCCCTTGGTATGATGACTG 540  
 Db 549 GCTGACAAAGGACAATTCCTGGAAGTGTGTAACAGGTTCCCTTGGTATGATGACTG 608  
 QY 541 GTGTTACAGTAGGAGGAGGAGTGTGCTGGAATAGCCATTACAAATTAAGCCCTG 600  
 Db 609 GTGTTACCTGTAGTGGAGGAGGAGTGTGCTGGAATAGCAATTTACAGTTAAGGCTG 668  
 QY 601 AAACCTAGTGTGAAGTATACGCTCTGAGCCCTGGAATGAGATGAGTCTACCAAGTCT 660  
 Db 669 AAACCTAGTGTGAAGTATATGCTGTAACCCCTCAATGAGATGAGTCTACCAAGTCT 728  
 QY 661 AAACCTGAAGGAGAACTGACCCCAATCTTCATCTCCAGAAACCATAGCAGATGGTCT 720  
 Db 729 AAGCTGAAGGAGAACTGATGCCAATCTTTATCTCCAGAAACCATAGCAGATGGTCT 788  
 QY 721 AAATCCAGATGGCTTCAATACCTGGCTTATATAGAGACCTTGTGGATGATGCTCTC 780  
 Db 789 AAATCCAGATGGCTTCAACACCTGGCTTATATAGAGACCTTGTGGATGATGCTCTC 848  
 QY 781 ACTGTACCCGAGATGAATCAAGTATGCAACCCAGCTGGTGGGGAGAGATGAACCTG 840  
 Db 849 ACTGTACAGAGGATGAATTAAGTGTCAACCCAGCTGGTGGGGAGAGATGAACCTA 908  
 QY 841 CTCAATTAGCCGACGCTGGCGTGCACATGCTGCGAGTGTGCTCAACATTTCCAAACA 900  
 Db 909 CTCAATTGAACCTACAGCTGGTGGTGGAGTGTGCTGCTCTCAACATTTTCAAACT 968  
 QY 901 GTCTCTCAGAGTAAAGACGCTGTCATTTACTACGTGGGGGGAATGTAGACCTAA-- 958  
 Db 969 GTTCCCCAGAGTAAAGAACATTTGTATGTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
 QY 959 -CCTCCCTGAACCTGGTGGGAGGCTGAACGGCCAGCTCTTACAGACAGCTGTGTTT 1016  
 Db 1029 TCCCTCAATACTGGGTGAAGCAGCTGAAGGCCAGCTTCTATCAGTCTGTTTCTGT 1087

## RESULT 7

AAI70575

ID AAI70575 standard; cDNA; 1336 BP.

AC AAI70575;

XX 21-JAN-2002 (first entry)

DE Human serine racemase cDNA.

XX Serine racemase; human; D-serine; regulation;

KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;

KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;

KW nootropic; neuroprotective; cerebroprotective; antiparkinsonian;

KW analgesic; diagnosis; gene therapy; screening; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers  
 FH CDS 52..1074  
 FT /\*tag= a  
 XX  
 PN WO200173077-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-EP03668.  
 XX  
 PR 31-MAR-2000; 2000US-193748P.  
 PR 03-APR-2000; 2000US-194249P.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Ramakrishnan S;  
 XX  
 DR WPI; 2001-648444/74.  
 DR P-PSDB; AAM50262.  
 XX  
 PS Polynucleotide encoding serine racemase enzyme and the enzyme useful  
 for screening reagents regulating the activity of the enzyme in a  
 neuron disease caused by over- or under-activation of glutamate  
 N-methyl-D-aspartate  
 Claim 1; Fig 1; 66pp; English.  
 XX  
 CC The present sequence is that of cDNA encoding human serine racemase  
 (see AAM50262). The polynucleotide can be used in the production  
 of recombinant serine racemase enzyme, or for the detection of  
 serine racemase polynucleotides. Expression vectors and host cells  
 are claimed. Serine racemase catalyses the conversion of L-serine  
 to D-serine. Neuron damage following various nervous system diseases  
 is often caused by activation of glutamate N-methyl-D-aspartate  
 (NMDA) receptors in the brain. This receptor is activated by the  
 binding of D-serine. Regulation of D-serine levels through  
 regulation of serine racemase may therefore prevent or minimise  
 neuron damage in neurogenic and myopathic disorders,  
 neurodegenerative disorders such as Alzheimer's disease and  
 Parkinson's disease, and disorders leading to peripheral and  
 chronic pain. Serine racemase polypeptides and polynucleotides are  
 used in claimed methods of screening for agents that modulate or  
 decrease the activity of serine racemase. Also claimed is a  
 pharmaceutical composition comprising either an expression vector  
 that contains a serine racemase polynucleotide, or a reagent that  
 modulates serine racemase enzyme activity. This is used to modulate  
 serine racemase activity in a disease, particularly neuron damage or  
 a neurodegenerative disease caused by the over- or under-activation  
 of the glutamate NMDA receptor.  
 XX  
 SQ Sequence 1336 BP; 386 A; 268 C; 319 G; 362 T; 1 other;

Query Match 79.2%; Score 806.2; DB 22; Length 1336;  
 Best Local Similarity 87.6%; Pred. No. 1.5e-246;  
 Matches 893; Conservative 0; Mismatches 123; Indels 3; Gaps 1;  
 QY 1 ATGTGTGCTCAGTACTGTCATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 60  
 Db 52 ATGTGTGCTCAGTATTGTCATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 111  
 QY 61 GACTCTATCCACCTCACCCAGTGTCTTCAACAGCTCCATTTTGAATCAATAGCAGGCGC 120  
 Db 112 GATTCTATCCACCTCACACAGTGTCTTCAACAGCTCCATTTTGAATCAACAGGCGC 171  
 QY 121 ATCTCTTTCTTCAATGTGAGCTCTTCCAGAAACTGGTCTTTTAAGATTTCAGGTGCC 180  
 Db 172 AATCTTTTCTTCAATGTGAGCTCTTCCAGAAACTGGTCTTTTAAGATTTCAGGTGCC 231  
 QY 181 CTTAATGCCATCAGAGGCTTAACTTCTGACACGCCAGAGAGAACCCCAAGCCCTAGTGT 240  
 Db 232 CTCATGCCGTCAGAGGCTTGGTTCTCTGATGCTTTTAGAAGAGGAGCCGAAAGCTGTGTT 291



QY 661 AAATCGAAGGAGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGC 720  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 922 AAGCTGAAGGGGAACTGATGCGCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 981  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 721 AAATCCAGATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTGTGGATGATCTTC 780  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 982 AAATCCAGATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTGTGGATGATCTTC 1041  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 781 ACTGTCCAGCAAGTGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAAACAG 840  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1042 ACTGTCCAGCAAGTGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAAACAG 1101  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 841 CTCATTGAGCCGACTGCTGGCGGTGGCACTGGCTGCAAGTGTCTCAGCATTTCCAAACA 900  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1102 CTCATTGAACTACAGCTGGTGTGGAGTGGCTGCTGCTCAACATTTTCAAACT 1161  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 901 GTCTCTCCAGCAAGTGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAAACAG 958  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1162 GTTCTCTCAGCAAGTGAATCAAGTATGCAACCCAGCTGGTGTGGGGAGAAATGAAACAG 1221  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 959 -CTCCTCTCAACTGGGTGGGGCAGGCTGAACGGCCAGCTCCTTACCAGACGGTCTGTTT 1016  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 1222 TCCTCCATACTTGGGTGAAGCAGGCTGAACGGCCAGCTCCTTACCAGACGGTCTGTTT 1280  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 RESULT 9  
 AAH06600  
 ID AAH06600 standard; cDNA; 848 BP.  
 XX  
 AC AAH06600;  
 XX  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:3435.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 27-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 3435; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH93893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 848 BP; 236 A; 187 C; 214 G; 207 T; 4 other;  
 Query Match 58.9%; Score 599.6; DB 22; Length 848;  
 Best Local Similarity 87.6%; Pred. No. 1.le-180;  
 Matches 687; Conservative 0; Mismatches 93; Indels 4; Gaps 3;  
 QY 1 ATGTGTGCTCAGTACTGCATCTCCTTTGCTGATGTTCAAAAAGCTCATATCAACATTCAA 60  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 69 ATGTGTGCTCAGTATTGTCATCTCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCAA 128  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 61 GACTCTATCCACCTCACCCCAAGCTTAACAAAGCTTCCATTTTGAATCAAAATAGCAGGCGC 120  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 129 GATTCTATCCACCTCACCCAGCTGCTAACAAGCTCCATTTTGAATCAAAATAGCAGGCGC 188  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 121 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAAAGCTGGTCTTTTAAAGATTCGAGGTGCC 180  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 189 AATCTTTTCTCAAAATGTGAACTCTTCCAGAAAAGCTGATCTTTTAAAGATTCGAGGTGCC 248  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 181 CTTAATGCCATCAGAGGCTTAATTCCTGCACACGCCAGAGAGAGCCCAAGCCGTAGTT 240  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 249 CTCATATGCCCTCAGAAGCTTGGTTCCTGATGCTTTTAAAGAAAGAGCCGAAGCTGTGTT 308  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 241 ACTCAGACAGCGGAAACCAATGGCCAAAGCTCTCACCTATGCTGCTAACTTGGAGGAATT 300  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 309 ACTCAGACAGCTGGAACCAATGGCCAGGCTCTCACCTATGCTGCTGCAAAATTTGGAAGGAATT 368  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 301 CTTGCTTACATTTGTTTCCCAACAGCTTCCCAAGTGCAGAAAGCTGGCAATCCCAAGCC 360  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 369 CTTGCTTATATTGTTGTCGCCCAAGAGCTGTAAGAAAGCTGTAAGAAAGCTGTAAGAAAG 428  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 361 TATGGACATCATAGTATATCTGTGACCCCAAGTGCAGAGTCCAGAGAAAAGGTCACTCAA 420  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 429 TACGGAGCGTCAATTTGTTATCTGTGAACCTAGTGTAGTCCAGAGAAAATGTTGCAAAA 488  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 421 AGAATTATGCAAGAAACAGAGGCTCTTGTGCTCCATCCCAAGAGGAGCTGAGTGATA 480  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 489 AGATTTACAGAAACAGAGAGGCTATGTTGATACATCCCAAGAGGAGCTGAGTGATA 548  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 481 GCTGGACAGGAAATTTGCCCTGGAAAGTGTCTGAACCAAGGTTCCTTTGGTAGATGCACTG 540  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 549 GCTGGACAGGAGCAATTTGCCCTGGAAAGTGTCTGAACCAAGGTTCCTTTGGTAGATGCACTG 608  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 541 GTGTTACCAGTAGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAGGCCCTG 600  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 609 GTGTTACCAGTAGGAGGAGGAGGAATGGTTGCTGGAAATAGCCATTACAATTAAGGCCCTG 668  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 601 AAACCTAGTGTGAAGGTATACGCTGTGAGCCCTCGAATCGATGATGCTCTACCACTCT 660  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 669 AAACCTAGTGTGAAGGTATATGCTGTGAGCCCTCAATCGATGATGCTG-TTCCAGTCC 727  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 661 AAATCAAGGAGAACTGACCCCAATCTTATCTCCAGAAACCATAGCAGATGGTGC 720  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 728 AAGCTGAAGGGGAACTGATGCGCAATCTTTATCTCCAGAAACCATAGCAGATGGTGC 785  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 QY 721 AAATCCAGATTTGGCTTGAATACCTGGCCCTATTATTAAGAGACCTTGTGGATGATGCTTC 780  
 || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 786 AAATCCA-CATTGGCTTGAACACCTGGGCTATTATTATCANGGACCTTGGGATGATATCTTT 844

Qy 781 ACTG 784  
 Db 845 ACTG 848

RESULT 10  
 ABL90122  
 ID ABL90122 standard; cDNA; 731 BP.  
 XX  
 AC ABL90122;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 684.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 XX 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PA Birse CE, Rosen CA;  
 XX  
 PI WI: 2002-122018/16.  
 XX  
 DR P-PSDB; ABB89713.  
 XX  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive, and  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 XX Claim 4; SEQ ID NO 684; 2081pp + Sequence Listing: English.  
 XX  
 CC The invention relates to novel genes (ABL9449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 731 BP; 204 A; 158 C; 184 G; 178 T; 7 other;

Query Match  
 Best Local Similarity 52.1%; Score 530.4; DB 24; Length 731;  
 Matches 598; Conservative 1; Mismatches 81; Indels 2; Gaps 2;

Qy 1 ATGTGTGCTCAGTATTCATCTCCTTTGGTGTGATGTTGAAAAAGCTCATATCAACATTCGA 106  
 |||||  
 Db 61 GACTCTATCCACCTCACCCAGTCTAACAGCTCCATTTTGAATCAATAGAGGCGC 120  
 |||||  
 Db 107 GATTCTATCCACCTCACCCAGTCTAACAGCTCCATTTTGAATCAATAGAGGCGC 166  
 |||||  
 Qy 121 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAGATTCGAGGTGCC 180  
 |||||  
 Db 167 AATCTTTTCTCAAAATGTGAGCTCTTCCAGAAACTGGGTCTTTTAAGATTCGAGGTGCC 226  
 |||||  
 Qy 181 CTTAATGCCATCAGAGCTTAAATCTTGACACGCCAGAGAGGCCAAAGCCGTAGTT 240  
 |||||  
 Db 227 CTCATGCCCTCAGAGCTTGGTTCCTGATGCTTTAGAAAGAGGCGAAAGCTGTGTT 286  
 |||||  
 Qy 241 ACTCAGACGAGCGAACCATGCCAGCTCTCACCTATGCTGCTAACTGGAGGAATTT 300  
 |||||  
 Db 287 ACTCAGACGAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAATTT 346  
 |||||  
 Qy 301 CCTGCTTACATTTGTTGTTCCCAACAGCTCTCCCAACTGCAAGAACTGGCAATCCCAAGCC 360  
 |||||  
 Db 347 CCTGCTTATATTGTTGTTGCCCAAGAGCTCCAGACTGTAAANAATTTGCAATACAAGCC 406  
 |||||  
 Qy 361 TATGGAGCTCATGATATATCTGTGACCCCAAGTGAGAGTCCAGAGAAAAGGCTCACTCAA 420  
 |||||  
 Db 407 TACGGAGCGCTCAATTTGTTATCTGTGAACCTAGTGTGATGAGTCCAGAGAAAATGTTGCAAAA 466  
 |||||  
 Qy 421 AGAATTATCAAGAAACAGAGGAGCATCTTGGTCCATCCCAACAGGAGGCTCGCATCATATA 480  
 |||||  
 Db 467 AGAGTTACAGAGAAAGAAACAGAGGAGCATCTTGGTACATCCCAACAGGAGGCTCGCATGATA 526  
 |||||  
 Qy 481 GCTGGACAGGAAACAATTTGCCCTGGAAGTCTCTGAACAGGTTCCCTTTGATGACACTG 540  
 |||||  
 Db 527 GCTGGACAGGAGCAATTTGCCCTGGAAGTCTCTGAACAGGTTCCCTTTGATGACACTG 586  
 |||||  
 Qy 541 GTGGTACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTAACAATTAGGCCCTG 600  
 |||||  
 Db 587 GTGGTACCCTGTRGGTGGAGGAGGAGGATGCTTGTGGAATAGCAATTAACAATTAGGCCCTG 646  
 |||||  
 Qy 601 AAACCTAGTGTGAAGTATACGCTGCTGAGCCCTCGAATCGCATGCTGCTACCTGCTACCTGCT 660  
 |||||  
 Db 647 AAACCTAGTGTGAAGTATATGCTGNTGA-NCCTCAATGCAGATGACTG-TACCAGTCC 704  
 |||||  
 Qy 661 AAACCTGAAGGAGAGACTGACCC 682  
 |||||  
 Db 705 AAGCTGAAGGGGACTTATGCCC 726

RESULT 11  
 AAA59295  
 ID AAA59295 standard; DNA; 608 BP.  
 XX  
 AC AAA59295;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE N-terminal sequence of human serine racemase DNA.  
 XX  
 KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
 KW neural dysfunction; NMDA receptor; Parkinson's disease;  
 KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000043526-A1.  
 XX  
 PD 27-JUL-2000.  
 XX  
 XX 18-JAN-2000; 2000WO-US00938.  
 XX  
 PR 19-JAN-1999; 99US-0116333.  
 PR 21-JUL-1999; 99US-0144839.  
 PR 28-JUL-1999; 99US-0145953.  
 XX

(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
 Ferris CD;  
 WPI; 2000-482915/42.  
 Mammalian serine racemase preparations, used to identify modulators  
 which can be used to treat diseases associated with  
 N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
 Claim 17; Page 26; 54pp; English.  
 The present sequence represents a fragment of a mammalian serine  
 racemase gene. The racemase polypeptide has a specific activity of at  
 least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the  
 direct racemisation of L-serine to D-serine. D-serine appears to be  
 an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The  
 mammalian serine racemases can be used to identify modulators, which  
 can be used in the treatment of acute or chronic neural death or  
 dysfunction mediated by overactivation of N-methyl-D-aspartate  
 receptors. Overactivation of the receptors is associated with  
 Parkinson's disease, Huntington's disease, motor neurone disease and  
 Alzheimer's disease.  
 Claim 18; Page 27; 54pp; English.  
 The present sequence represents a fragment of a mammalian serine  
 racemase gene. The racemase polypeptide has a specific activity of at  
 least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the  
 direct racemisation of L-serine to D-serine. D-serine appears to be  
 an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The  
 mammalian serine racemases can be used to identify modulators, which  
 can be used in the treatment of acute or chronic neural death or  
 dysfunction mediated by overactivation of N-methyl-D-aspartate  
 receptors. Overactivation of the receptors is associated with  
 Parkinson's disease, Huntington's disease, motor neurone disease and  
 Alzheimer's disease.



PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-541565/60.  
XX

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure; SEQ ID NO 8040; 1701pp + Sequence Listing; English.

PS The invention relates to novel genes (AB11004-ABA21534) and proteins (AB114678-AB118001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

XX Sequence 20892 BP; 5768 A; 4125 C; 5164 G; 5835 T; 0 other;

Query Match 17.9%; Score 181.8; DB 22; Length 20892;  
Best Local Similarity 88.0%; Pred. No. 1.7e-46;  
Matches 198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 593 AGGCCCTCAACCTAGTGTGAGGTATACGCTGCTGACCCCTCGAATCGACATGCTCT 652  
DB 13313 AGGCTCTCAACCTAGTGTGAGGTATATGCTGTGACCCCTCAATCGACATGCTCT 13254  
QY 653 ACCAGTCTAAACTGAAGAGGAAGTACGCCCAATCTTCATCTCCAGAAACCATAGCAG 712  
DB 13253 ACCAGTCAACCTAGTGAAGGGAACATGATGCCCAATCTTCATCTCCAGAAACCATAGCAG 13194  
QY 713 ATGGTGTCAAAATCCAGCATTCGCTTGAATACCTGGCCTTATTATAAGAGACCTTGTGGATG 772

Db 13193 ATGGTGTCAAAATCCAGCATTCGCTTGAACACCTGCTTATTATCAGGACCTGTGGATG 13134  
QY 773 ATGCTCTTCACTGTCTACCCGAAGATGAATCAAGTATGCAACCCAGC 817  
Db 13133 ATATCTTCACTGTCTACAGAGGATGAATTAAGGTGAGGCTCCAGC 13089

#### RESULT 14

AAAC10869  
ID AAC10869 standard; cDNA; 861 BP.

XX  
AC AAC10869;

XX  
DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 14944.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EPI033401-A2.

XX  
PD 06-SEP-2000.

XX  
PF 21-FEB-2000; 2000EP-0200610.

XX  
PR 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 14944; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX Sequence 861 BP; 257 A; 167 C; 240 G; 191 T; 6 other;

Query Match 16.5%; Score 168.2; DB 21; Length 861;  
Best Local Similarity 88.8%; Pred. No. 6.6e-43;  
Matches 182; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 394 GACGAGTCCAGAGAAAGGTCACCTCAAGACATTATGCAAGAACAGAGGATCTGTGTC 453  
DB 657 GAGATGTCCAGAGAAATGTTGCAAAAAGAGTTACAGAGAAACAGAGGATCATGTGTA 716  
QY 454 CATCCCAACCAAGGAGCCTGCTGAGTGTAGCTGGACAAAGGAACAATTCCTTGAAGTGTG 513  
DB 717 CATCCCAACCAAGGAGCCTGCTGAGTGTAGCTGGACAAAGGAACAATTCCTTGAAGTGTG 776  
QY 514 AACCAAGGTTCCCTTGGTGTAGTGCATGCTGTGTACCAAGTAGGAGGAGGAATGTTGCT 573

```
Db 777 AACCAGGTTCTTGGTGGATCCACTGGTGTACCTGTAGGTGGAGGAATGCTGCT 836
QY 574 GGAATAGCATTACAAATTAAGGCC 598
Db 837 GGAATAGCATTACAGTTAAGGCTC 861

RESULT 15
AAV74370
ID AAV74370 standard; DNA; 31096 BP.
XX AC AAV74370;
XX 16-MAR-1999 (first entry)
XX Staphylococcus aureus contig SEQ ID #59.
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 1201..1260
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3001..3060
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 4801..4860
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 6601..6660
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 8401..8460
FT /*tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 10201..10250
FT /*tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 12001..12060
FT /*tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 13801..13860
FT /*tag= h
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"

FT misc_feature 15601..15660
FT /*tag= i
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 17401..17460
FT /*tag= j
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 19201..19260
FT /*tag= k
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 21001..21060
FT /*tag= l
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 22801..22860
FT /*tag= m
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 24601..24660
FT /*tag= n
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 26401..26460
FT /*tag= o
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 28201..28260
FT /*tag= p
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 30001..30060
FT /*tag= q
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX EP786519-A2.
XX 30-JUL-1997.
XX 07-JAN-1997; 97EP-0100117.
XX 05-JAN-1996; 96US-0009861.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX WPI; 1997-374922/35.
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT stored on computer readable medium and used in the production of
```



anti-S.aureus vaccines

Claim 1; Page 452-469; 327lpp; English.

This sequence represents one of 5191 Staphylococcus aureus DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium.

Sequence 31096 BP; 11857 A; 5243 C; 5477 G; 7488 T; 1031 other;

Query Match 13.7%; Score 139.8; DB 18; Length 31096;  
Best Local Similarity 48.5%; Pred. No. 5.4e-33;  
Matches 450; Conservative 0; Mismatches 468; Indels 10; Gaps 2;

Qy	35	TTGAAAAGCTCATATCAACATTCAGACATCTATCCACCTCACCCAGTGTACAAAGCT 94
Db	1261	TCGAAGAAGCTTAAAGCAAGCATTAACACATTTATTCGTGCAACACCTCTAATTAATCAAA 1320
Qy	95	CCATTTTGAATCAA---ATAGCAGGGCGCAATCTTTCTTCAAAATGTGAGCTCTCCAGA 151
Db	1321	TGTATTTAAGCCAAAGTATACTAAAGGAATGTATTTCTAAAAATGAGAAATATGCAAT 1380
Qy	152	AAATCGGTCTTTTAAAGATTCGAGGTGCCCTTAATGCCATCAGAGGCTTAATTCCTGACA 211
Db	1381	TCACAGGATCTTTTAAATTTAGAGCGCTAGCAAT-----NAAATTAATCACTTAAC 1433
Qy	212	CGCCAGAAGAGACCCCAAGCCGTAGTTACTACAGCAGCGGGAACCATGGCCAAAGCTC 271
Db	1434	AGATGAACAAAGAAAGAAAGGCATTTATCGCAGCATCTGCTGGGGAAGCCATGCACAAGGTG 1493
Qy	272	TCACCTATGCTGTAAACTTGAAGGAATTCCTGCTTACATTTGCTGCTCCCAACAGCTC 331
Db	1494	TTGCTTTAAGCAGTAAATTTATAGGCATTTGATGCAACGATTTGTAATGCCCTGAACACGAC 1553
Qy	332	CCAACTGCAAGAACTGGCAATCCAAAGCCTATGGAGCATCGATAGTATACTGTGACCCAA 391
Db	1554	CACAAGCGAAACAACAAGCAACAAAGGCTATGGGCAAGGTTATTTTAAAGGTAATAA 1613
Qy	392	GTGACGAGTCCAGAGAAAAGGTCACTCAAGAAATTTGCAAGAAACAAGAGGCATCTTGG 451
Db	1614	ACTTTAAGAACTAGACTTTATATGGAAGAATTTAGCGAAAGAAAATGGCATGACAATCG 1673
Qy	452	TCCATCCCAACAGGAGCCCTGAGTGATAGCTGGACAAGGAACAATTCCTCGGAAGTGC 511
Db	1674	TTCATCCATATGAGGATAAGTTTGTATGGCAGCCCAAGNACAATTTGGTTAGAAATTT 1733
Qy	512	TGAACAGGTTCCCTTGGTAGATGCATCGTGTGTGTTACAGTAGGAGGAGGAATGGTTG 571
Db	1734	TAGATGATATTTGGAATGTGAATACAGTCATCGTACCAGTTGGCGGTGGAGGATTAATTG 1793
Qy	572	CTGGAATAGCCATTAACAATTAAGCCCTGAAACCTAGTGTGAAGGTATACGCTCTGAGC 631
Db	1794	CAGGTATGGCCAGCCATTAATAATCATTTAAACCTTCAATTTATTCGGTGTTCAT 1853
Qy	632	CCTCGAATGCAATGACTGTCTACCAAGTCTAAACTGAAAGGAGAACTGACCCCAATCTTC 691
Db	1854	CTGAAATGTTATGGTATGGCTGAGTCTTTCTTATAGAGAGATTTAACTGAACATCGAG 1913

Search completed: June 24, 2003, 04:26:07  
Job time : 291 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 03:22:18 ; Search time 186 Seconds  
(without alignments)  
8031.403 Million cell updates/sec

Title: us-09-889-609b-1

Perfect score: 1018

Sequence: 1 atgtgtgtcagtcactgcatt.....ttaccagacggtctgttttaa 1018

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.4	14.1	31096	7	US-08-781-986A-59
2	87.4	8.6	1830121	9	US-10-329-960-1
3	77	7.6	954	10	US-09-974-300-4749
4	71	7.0	105184	9	US-09-847-513A-1
5	70.2	6.9	930	9	US-09-738-626-1091
6	70.2	6.9	3309400	9	US-09-738-626-1091
7	60.6	6.0	1545	9	US-09-942-891-1
8	60.6	6.0	1545	9	US-09-942-891-5
9	60.6	6.0	1545	9	US-09-942-891-7
10	60.6	6.0	1545	9	US-09-942-891-8
11	55.2	5.4	1251	10	US-09-974-300-264
12	52.6	5.2	536165	9	US-09-939-964-1
13	46	4.5	926	10	US-09-974-300-4809
14	38.8	3.8	1379	9	US-10-098-841-197
15	38.4	3.8	15249	7	US-08-781-986A-102
16	37.2	3.7	1744	10	US-09-804-682-78
17	36.8	3.6	6320	9	US-10-037-270-135
18	35.2	3.5	1061	10	US-09-833-381-1809
19	34.6	3.4	31871	9	US-10-092-154-1403

20	34.6	3.4	31871	10	US-09-764-847-1403	Sequence 1403, Ap
21	34.6	3.4	44848	9	US-09-988-113-42	Sequence 42, Appl
22	34.6	3.4	44848	10	US-09-776-874A-42	Sequence 42, Appl
23	34.4	3.4	1049	9	US-10-123-155-358	Sequence 358, App
24	34.2	3.4	928	10	US-09-974-300-271	Sequence 271, App
25	34	3.3	486	9	US-09-918-995-32623	Sequence 32623, A
26	34	3.3	3655	10	US-09-903-187A-6	Sequence 6, Appli
27	34	3.3	3657	9	US-09-903-170C-6	Sequence 6, Appli
28	34	3.3	3657	10	US-09-903-180B-6	Sequence 6, Appli
29	34	3.3	3657	10	US-09-903-171A-6	Sequence 6, Appli
30	34	3.3	3657	10	US-09-903-188A-6	Sequence 6, Appli
31	34	3.3	3657	10	US-09-903-323A-6	Sequence 6, Appli
32	34	3.3	3657	10	US-09-903-325A-6	Sequence 6, Appli
33	33.8	3.3	498	10	US-09-864-761-32574	Sequence 32574, A
34	33.8	3.3	1274	9	US-09-983-802-75	Sequence 75, Appl
35	33.6	3.3	677	9	US-10-123-155-230	Sequence 230, App
36	33.4	3.3	1569	10	US-09-974-300-2651	Sequence 2651, Ap
37	33.4	3.3	2698	10	US-09-939-980-207	Sequence 207, App
38	33	3.2	1827	10	US-09-974-300-1928	Sequence 1928, Ap
39	33	3.2	6459	9	US-10-029-217A-6	Sequence 6, Appli
40	33	3.2	81940	9	US-09-759-508B-1	Sequence 1, Appli
41	32.8	3.2	552	9	US-10-123-155-196	Sequence 196, App
42	32.8	3.2	569	9	US-10-066-543-522	Sequence 522, App
43	32.8	3.2	1956	10	US-09-351-794A-1	Sequence 1, Appli
44	32.6	3.2	398	10	US-09-880-107-881	Sequence 881, App
45	32.6	3.2	763	9	US-10-153-668-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1  
US-08-781-986A-59  
Sequence 59, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31096 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-781-986A-59

Query Match		14.1%	Score 143.4;	DB 7;	Length 31096;
Best Local Similarity		48.6%	Pred. No. 1.1e-37;		
Matches 456;		Conservative 0;	Mismatches 472;	Indels 10;	Gaps 2;
QY	25	TTTGCTGATGTTCAAAAGCTCATATCAACATTCAGACTCTATCCACCTCACCACGAGTG	84		
DB	1251	TTAGGAGATGATGAAGAGCTAAAGCAAGCATTAACACCATTTATTGTCGCAACACCTCTA	1310		
QY	85	CTAACAGCTCCATTTTGAATCAA---ATAGCAGGGCGCAATCTTTTCTTCAAAATGTGAG	141		
DB	1311	ATTAATCAATGATTTAAGCCAAAGTATACTAAAGGAATGTATTTCTAAATATAGAA	1370		
QY	142	CTCTCCAGAAAGCTGGGCTCTTTTAAGATTCGAGTGGCCCTTAATGCCATCAGAGCTTGA	201		
DB	1371	AATATGCAATTCACAGGATCTTTTAAATTTAGAGCGCTAGCAAT-----NAAATTA	1423		
QY	202	ATTCCTGACAGCCAGAGAGAGCCCAAGCGGTAGTTACTCAGCAGCAGCGGAACCAAT	261		
DB	1424	ATCACTTAACAGATGAACAAAAGAAAAGGCAATTATCGCAGCATCTGCTGGGGAACCAT	1483		
QY	262	GGCCAAGCTCTCACCTATCTGCTAACTGGAAGGAATTCCTGCTTACATTTGTGGTCC	321		
DB	1484	GCACAGGTGTGCTTTACACGCTAAATTTATAGGCATTTGATGCACAGATTTGATGCT	1543		
QY	322	CAACAGCTCCCACTGCAAGAACTGGCAATCCAAGCCTATGGAGCATCGATAGTATAC	381		
DB	1544	GAACAGCACCACAGCGGAAACAAACAACGCAAAAAGGCTATGGGCAAAAGGTTATTTA	1603		
QY	382	TGTGACCCCACTGACGAGTCCAGAGNAAGGTCACCAAGAAATATGCAAGAACAGAA	441		
DB	1604	AAAGSTAAAACCTTTAAACGAACTAGACTTTATATGGAAGAATTAGCGAAAGAAATGCG	1663		
QY	442	GGCATCTTGGTCCATCCCAACAGAGCGCTGCAGTGATAGCTGGCAAGGAACAATGGC	501		
DB	1664	ATGACAACTGTCATCCATATGACGATAAGTTTGTATGGCAGGCCAAGGAACAATGTT	1723		
QY	502	CTGGAAGTCTGAAACAGGTTCCCTTGGTAGATGACACTGGTGGTACCAGTAGGAGAGA	561		
DB	1724	TTAGAAATTTAGATGATATTTTGAATGTGAATACAGTCATCGTACCAGTTGGCGGTGA	1783		
QY	562	GGAATGGTTGGTGGATACCATTAACATTAAGGCCCTGAAACCTAGTGTGAAGGTATAC	621		
DB	1784	GGATTAATTTGCAGGTATTGCCACCGCATTAATAATCATTTAAACCTTCAATTCATATATC	1843		
QY	622	GCTGCTGAGCCCTCGAATCGATGACTGCTACCAGTCTAAACTGAAAGGAGAAGCTGACC	681		
DB	1844	GGTGTCAATCTGAGAAATGTTTATGATGGCTGAGTCTTTCTATAAGAGAGATTTAACT	1903		
QY	682	CCCAATCTTCATCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCAATTCGCTTGAAT	741		
DB	1904	GAACATCGAGTGGATAGCAATAGCAGATGGTGTGATGTAAAGTTCCTGCTGAACAA	1963		
QY	742	ACCTGGCTATTATAGAGACCTTGGGATGATGCTTTCACCTGTCCACCAAGATGAATC	801		
DB	1964	ACATATGAAGTAGTTAAACATTTAGTAGATGAATTTATTTCTTTACTGAAGAAGAAAT	2023		
QY	802	AAATATGCAACCACTGGTGTGGGGAGAATGAAACTGCTCATTTGAGCCGACTGCTGCG	861		
DB	2024	GAACATGCTATGAAGATTTAATGCAGCTGCGCAAAATTTACTCAAGTTCAGCGCA	2083		
QY	862	GTGGCACTGGCTGCAGTGTCTGCAGCAATTTCCAAACAGTCTCTCCAGAACTAAGAAC	921		
DB	2084	TTACCAACAGCTGCAATTTTAAAGTGGAAAAATAAACAATAAATGGCTTGAAGATAAAAT	2143		
QY	922	GTCTGCTGCTACTCAGTGGGGGAATGTAGACCTAAC	959		
DB	2144	GTTGTTGCATTTAGTTTCAGCGGGGAATGTTGACTTAAC	2181		

RESULT 2  
US-10-329-960-1  
; Sequence 1, Application US/10329960  
; Publication No. US2003009277A1

GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F  
; FILE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB186P1  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40808)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature

```
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Query Match 8.6%; Score 87.4; DB 9; Length 1830121;  
Best Local Similarity 45.7%; Pred. No. 3.9e-17;  
Matches 304; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

```
QY 215 CAGAGAGAGCCCAAGCGGTAGTTACTACAGACGCGGAAACCACTGGCCCAAGCTCTCA 274
      ||||| ||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794187 CAGACAAAAAGCAGCTGGCGTAATAGCAGCCCTCTCGGGGTAAACATGCGCAAGCGGTGG 794246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 275 CCTATGCTGCTAAACTGGAAGGAATTCCTGCTTACATTTGGTTGCCCAACAGCTCCCA 334
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794247 CATATCGCGGAAACAATTAGGCTTAAAGCATTAATTTATGCCACAAAACACCCCAA 794306
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 335 ACTGCAAGAACTGGCAATCCAAAGCCTATGGAGCATCGATAGTATATCTGTGACCCCAAGTG 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794307 GCATTAAGTGGATGCAGTCGCGTGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 794366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 ACGAGTCCAGAGAAAGGTCACCTCAAGAAATTTATGCAAGAAACAGAGGCACTTTGGTCC 454
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794367 TCGATGAAGCCAAAGCAAGGCAATCGAGCTTTCAAAGAAAAAAGACATGACATTTATTC 794426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



167	Db	ACGTGCGAGCTTCGGAACACGCACTACTCGACCAACGCGTTGGCATCGTCGCGGCATCAG	226
251	QY	GGGGAACCATGGCCAAAGCTCTCACTATGCTGTCTAAACTGGAAGAAATTCCTGCTTACA	310
227	Db	CGGGAACGCGAGGACTCGCAAAATGCTTTTGCCGCGAGCATCCTTAAGCGTTCCCGCCACGG	286
311	QY	TTGTGGTTCCCAACACAGCTCCCACTCCAGAAACACTGGGCAATCCAAAGCCTATGGAGCAT	370
287	Db	TATTGTGTCGCCGAAACTGCCCAACAAGTAAAGATTGATCGCCTCAAGCAATACGGTGCAA	346
371	QY	CGATAGTATACTGTGACCCCAAGTGACGAGTCCAGAGAAAGGTGACCTCAAGAAATATATGC	430
347	Db	CCGTGCAACAAATCCGGATCTGAATATGCGGAAGCATTTGAGGACGAGCTCAAAACCTTTGAGT	406
431	QY	AAGAAACAGAAAGGCATCTTTGGTCCATCCCAACAGGAGCGCTGCGAGTCATAGCTGGACAAG	490
407	Db	CGGAACCTGGTGCTCTGTTTTGCCACGCGCTACGACCGCCGACATCGCAGCTGGAGCAG	466
491	QY	GAACAAATTTGCCCTGGAAGTGTGTAACCAAGTTCCCTGGTAGATGCACTGGTGGTACCAG	550
467	Db	GCGTATTGGGCTAGAAATTCGAAGATCTTCCGACGTTGACACCATCTGCTGGTCTG	526
551	QY	TAGAGGAGGAGGAATGTTGCTGGAATAGC	581
527	Db	TCGGTTGGCGGTGCACTCTATCAGGAATCGC	557

## RESULT 6

```

US-09-738-626-1/C
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

	Query Match	Best Local Similarity	6.9%;	Score 70.2;	DB 9;	Length 3309400;
	Matches 213;	Conservative	0;	Mismatches 238;	Indels	Gaps 0
QY	131	TCAATGTGAGCTCTTCCAGAAAAC	TGGTCTTTTAAGATTTCGAGGTG	CGCCCTTAATGCCA	190	
Db	1039544	TCAAAGCAGAGTTCTTCCAAAAGT	CGCGCGTGTTCAAAACGCGTGGAGCAT	TTCACCGCC	103	
QY	191	TCAGAGGCTTAATTCTTCACGCGC	CAGAAGAGAGCCCAAGCGGTAGTT	TACTCACAGCA	250	
Db	1039484	AGCTCGCAGCTTCGGAACAGGAT	TACTCGACCCACGGTTGGCATCT	CGTCGGGCATCAG	103	
QY	251	GGCGAAACCATGCCCAAGCTCTC	ACCTATGCTGCTTAAACTGGAAGGAAT	TCTGCTTACA	310	

Db 1039424 GCGGAAACGAGGACTCGCAAAATGCTTTTGGCCGACGATCCTTTAAGCGTTTCCCGCACGG 1039365  
QY 311 TTCTGCTTCCCAACAGCTCCCAACTGCAAGAAACTGCGCAATCCAAAGCTTATGGAGCAT 370  
Db 1039364 TATTGTGGCCGAAACTGCCCAACAGTAAAGTTGATCGCCTCAAGCAATAGCGGTGCAA 1039305  
QY 371 CGATAGTATCTGTGACCCAAAGTGACGAGTCCAGAGAAAGGTCACCTCAAGAAATATGCG 430  
Db 1039304 CCGTGCACAAATCGGATCTGAATATGCGGAACATTTGAGGCGAGCTCAAACTTTGAGT 1039245  
QY 431 AGAAACACAGAGCATCTTTGGTCCATCCCAACAGAGCTGCGAGTATAGCTGGACAAG 490  
Db 1039244 CGGAAACTGTGTCTGTGTTTGGCCAGCTTACGACAGCGGACATCGCAGCTGGAGCAG 1039185  
QY 491 GAACAATTCCTCGGAGTGTCTGAACACAGGTTCCCTTGGTAGATGCACTGGTGGTACCAG 550  
Db 1039184 GCGTCAATGGCTGAGAAATTTGCGAAGATCTTCCCGACGTTGACACCATCGTGGTTGCTG 1039125  
QY 551 TAGGAGGAGGAGGAATGGTGTGCGAATAGC 581  
Db 1039124 TCGGTGCGGTGCACTCTATGCAGGAATGCG 1039094

## RESULT 7

US-09-942-891-1  
; Sequence 1, Application US/09942891  
; Publication No. US20030028917A1  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth James  
; APPLICANT: Mitsky, Timothy Albert  
; APPLICANT: Kishore, Ganesh Murthy  
; APPLICANT: Slater, Steven Charles  
; APPLICANT: Padgett, Stephen Rogers  
; APPLICANT: Stark, David Martin  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and  
; FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)  
; CURRENT APPLICATION NUMBER: US/09/942,891  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: US 09/313,123  
; PRIOR FILING DATE: 1999-05-17  
; PRIOR APPLICATION NUMBER: US 08/673,388  
; PRIOR FILING DATE: 1996-06-28  
; PRIOR APPLICATION NUMBER: US 08/628,039  
; PRIOR FILING DATE: 1996-04-04  
; PRIOR APPLICATION NUMBER: US 08/614,877  
; PRIOR FILING DATE: 1996-03-13  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1545  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-942-891-1

Query Match 6.0%; Score 60.6; DB 9; Length 1545;  
Best Local Similarity 46.5%; Pred. No. 7.9e-10;  
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 211 ACCCCAGAGAGAGAGCCCAAGCCGTAGTTACTACAGCAGCGGAAACCATGCCCAAGCT 270  
Db 220 ACGGAAGACAGAAAGCGCACGCGGTGATCACTGCTTCTCGGGTAAACCAACGCGCAGGCG 279  
QY 271 CTCACCTATGCTGCTAACTGGAAGAAATTCCTGCTTACATTTGTTGTTCCCAACAGCT 330  
Db 280 GTGCGGCTTTCTCTCGCGGTTAGCGGTGAAGCGCCTGATGCTGTTATGCAACCGCCACC 339  
QY 331 CCCAACTGCAAGAAACTGCAATCCAAGCTTATGGAGCATCGATAGTATATCTGTACCA 390  
Db 340 GCGGACATCAAGTCAAGCGGCTTACCGGCTTTCGGGGGGAAGTCTGCTTCCACGCGCG 399  
QY 391 AGTGACGATCCAGAGAAAGGTCACTCAAGAAATTTATGCAAGAAACAGAGGCAATCTTTG 450

Db 400 AACTTTGATGAAGCGAAACGCAAGCGATGCAACATGTCACAGCAGCGGGTTTCACTCGG 459  
QY 451 GTCCATCCCAACAGGAGCGCTGCAAGTATAGCTGGAACAGGACAATTTCCCTTGAAGTG 510  
Db 460 GTGCCGCCCTTCGACCATCCGATGGTGTATGCGGGCAAGGACGCTGCGCTGGAACGTG 519  
QY 511 CTGAACACAGGTTCCCTTGTGTAGATGCACTGGTGTACCACTAGGAGGAGGAGGAATGGTT 570  
Db 520 CTCAGCAGAGCGCCATCTCACCGCGTATTTGTGCCAGTGGCGGCGCGTCTGGCT 579  
QY 571 GCTGGAATAGCCATTACAATTAAGGCCCTGAAACCTAGTGTGAAAGTATACGCTGCTGA 629  
Db 580 GCTTGGCTGCGCGTGTGATCAACAACTGATGCCCAAACTCAAAAGTATGTCGCCGTAGA 638

## RESULT 8

US-09-942-891-5  
; Sequence 5, Application US/09942891  
; Publication No. US20030028917A1  
; GENERAL INFORMATION:  
; APPLICANT: Gruys, Kenneth James  
; APPLICANT: Mitsky, Timothy Albert  
; APPLICANT: Kishore, Ganesh Murthy  
; APPLICANT: Slater, Steven Charles  
; APPLICANT: Padgett, Stephen Rogers  
; APPLICANT: Stark, David Martin  
; TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and  
; FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)  
; CURRENT APPLICATION NUMBER: US/09/942,891  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: US 09/313,123  
; PRIOR FILING DATE: 1999-05-17  
; PRIOR APPLICATION NUMBER: US 08/673,388  
; PRIOR FILING DATE: 1996-06-28  
; PRIOR APPLICATION NUMBER: US 08/628,039  
; PRIOR FILING DATE: 1996-04-04  
; PRIOR APPLICATION NUMBER: US 08/614,877  
; PRIOR FILING DATE: 1996-03-13  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 1545  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-942-891-5

Query Match 6.0%; Score 60.6; DB 9; Length 1545;  
Best Local Similarity 46.5%; Pred. No. 7.9e-10;  
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 211 ACCCCAGAGAGAGAGCCCAAGCCGTAGTTACTACAGCAGCGGAAACCATGCCCAAGCT 270  
Db 220 ACGGAAGACAGAAAGCGCACGCGGTGATCACTGCTTCTCGGGTAAACCAACGCGCAGGCG 279  
QY 271 CTCACCTATGCTGCTAACTGGAAGAAATTCCTGCTTACATTTGTTGTTCCCAACAGCT 330  
Db 280 GTGCGGCTTTCTCTCGCGGTTAGCGGTGAAGCGCCTGATGCTGTTATGCAACCGCCACC 339  
QY 331 CCCAACTGCAAGAAACTGCAATCCAAGCTTATGGAGCATCGATAGTATATCTGTACCA 390  
Db 340 GCGGACATCAAGTCAAGCGGCTTACCGGCTTTCGGGGGGAAGTCTGCTTCCACGCGCG 399  
QY 391 AGTGACGATCCAGAGAAAGGTCACTCAAGAAATTTATGCAAGAAACAGAGGCAATCTTTG 450  
Db 400 AACTTTGATGAAGCGAAACGCAAGCGATGCAACATGTCACAGCAGCGGGTTTCACTCGG 459  
QY 451 GTCCATCCCAACAGGAGCGCTGCAAGTATAGCTGGAACAGGACAATTTCCCTTGAAGTG 510  
Db 460 GTGCCGCCCTTCGACCATCCGATGGTGTATGCGGGCAAGGACGCTGCGCTGGAACGTG 519



511	CTGACCAAGGTTCCCTTGGTAGTCACCTGGTGTACCACTAGAGAGAGAGAAATGGTT	570
Qy		
520	CTCCAGCAGGACGCCCATCTCGACGGCGTATTTGTGCACGTGCGCGCGCGCTTGGCT	579
Db		
571	GCTGGAATAGCCATTACAATTAAAGCCCTGAAACCTAGTGTGAAGTATACGCTGCTGA	629
Qy		
580	GCTTGGTGGCGGTGCTGATCAAAACACTGATGCCCAATCAAAAGTATAGCCGCTAGA	638
Db		

## RESULT 9

```

US-09-942-891-7
: Sequence 7, Application US/09942891
: Publication No. US20030028917A1
: GENERAL INFORMATION:
: APPLICANT: Gruys, Kenneth James
: APPLICANT: Mitsky, Timothy Albert
: APPLICANT: Kishore, Ganesh Murthy
: APPLICANT: Slater, Steven Charles
: APPLICANT: Padgett, Stephen Rogers
: APPLICANT: Stark, David Martin
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plants
: FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)
: CURRENT APPLICATION NUMBER: US/09/942,891
: CURRENT FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: US 09/313,123
: PRIOR FILING DATE: 1999-05-17
: PRIOR APPLICATION NUMBER: US 08/673,388
: PRIOR FILING DATE: 1996-06-28
: PRIOR APPLICATION NUMBER: US 08/628,039
: PRIOR FILING DATE: 1996-04-04
: PRIOR APPLICATION NUMBER: US 08/614,877
: PRIOR FILING DATE: 1996-03-13
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 7
: LENGTH: 1545
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
US-09-942-891-7

```

Query Match 6.0%; Score 60.6; DB 9; Length 1545;  
Best Local Similarity 46.5%; Pred. No. 7.9e-10;  
Matches 195; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

[illegible]

## RESULT 11

US-09-974-300-264

Db  
580 GCTTGGGTGGCGGTGCTGATCAACAACACTGATGCCGCAATCAAAGTGATCGCCGTAGA 638

## RESULT 10

```

RESOUR 10
US-09-942-891-8
: Sequence 8, Application US/09942891
: Publication No. US20030028917A1
: GENERAL INFORMATION:
: APPLICANT: Gruys, Kenneth James
: APPLICANT: Mitsky, Timothy Albert
: APPLICANT: Kishore, Ganesh Murthy
: APPLICANT: Slater, Steven Charles
: APPLICANT: Padgett, Stephen Rogers
: APPLICANT: Stark, David Martin
: TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly
: TITLE OF INVENTION: hydroxybutyrate-co-poly-beta-hydroxyvalerate in bacteria and
: FILE REFERENCE: 11899.0155.DVUS02 (MOBT:155--3)
: CURRENT APPLICATION NUMBER: US/09/942,891
: CURRENT FILING DATE: 2002-02-01
: PRIOR APPLICATION NUMBER: US 09/313,123
: PRIOR FILING DATE: 1999-05-17
: PRIOR APPLICATION NUMBER: US 08/673,388
: PRIOR FILING DATE: 1996-06-28
: PRIOR APPLICATION NUMBER: US 08/628,039
: PRIOR FILING DATE: 1996-04-04
: PRIOR APPLICATION NUMBER: US 08/614,877
: PRIOR FILING DATE: 1996-03-13
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8
: LENGTH: 1545
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
US-09-942-891-8

```

Query Match	6.0%;	Score 60.6;	DB 9;	Length 1545;
Best Local Similarity	46.5%;	pred. No. 7.9e-10;		
Matches 195; Conservative	0;	Mismatches 224;	Indels 0	

Qy		211	ACGCCAGAAGAGAACCCAAAGCCGTAGTTACTCACAGCAGCGGAACCAATGGCCAAGT	270
Db		220	ACGGAAGAACAAAGAGCAGCGGTGATCACTGCTTCGCGGTAAACCACGCAGGGC	279
Qy		271	CTCACCTATGCTGCTAACTGAAGGAATTCTGCTTACATTGTGGTTCCTCCCAAACAGT	330
Db		280	GTCGCGTTTTCTTCGCGCGTTAGGCGTGTAAGGCCCTGATCGTTATGCCAACC GCCCAC	339
Qy		331	CCC AACTCGCAAGAACTGGCAATCCAAGCCATATGGAGCATCGATAGTATAC TGTACCCA	390
Db		340	GCCGACATCAAGTCGACCGGCTGCGCGCTTCGGCGGGAAGTGCTGCTCCACGGCGG	399
Qy		391	AGTGACGAGTCCAGAGAAAAGGTCACTCAAAGAAATTATGCAAGAAACAAGAGGCATCTTG	450
Db		400	AAC TTGTATGAAGCGAAACGCAAGCCGATCGAACTGTCA CAGCAGCGGGTTCACTGG	459
Qy		451	GTCCATCCCAACACGAGGAGCTGCAGTGATAGCTGCACAAGGAACAATTGCCCTGGAAGTG	510
Db		460	GTGCGCGGTTCGACCATCCGATGTGTGATTCGGGGCAAGGACCGCTGCGCTGGAAC TG	519
Qy		511	CTGAACCAAGGTTCCCTTGGTAGATGCAC TGGTGTPACCA GTAGGAGGAGGAATGGTT	570
Db		520	CTCCAGCAGGACGCCATCTCGACCGGTATTTGTGCCAGTCGGCGCGCGCTGGCT	579
Qy		571	GCTGGATATAGCCATTACAATTAAAGCCCTGAA CCTAGTGTGAAGGATATAC GCTGCTGA	629
Db		580	GTTTCGTCGGCGGTGCTGATCAAAACA ACTGATGCCGCAAAATCAAAGTATATCCCGTAGA	638

## RESULT 11

US-09-974-300-264

```
; Sequence 264, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-264

Query Match          5.4%; Score 55.2; DB 10; Length 1251;
Best Local Similarity 44.9%; Pred. No. 5.2e-08;
Matches 354; Conservative 0; Mismatches 413; Indels 21; Gaps 3;

QY 18 CATCTCTTGTGCTGATGTTGAAAGAGCTCATATCAACATTCAGAGCTCTATCCACCTCAC 77
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 14 CATCCAAAGTCAAGACATTTTAAAGGCCCAACCAATGATAAAGATGTTGTTATTATCATAC 73
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 CCCAGTGCCTAACAGCTCCATTTTGAATCAAAATAGCAGGCGCAATCTTTTCTTCAAAATG 137
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 74 CCCTCTCCAAAAAATGAGAGGCTGTCGAGAGGTACGATTGCAAGCTATATTAAAAAG 133
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 138 TGAGTCTTCCAGAAACTGGGTCTTTTAAAGATTCGAGGTGCGCTTAATGCCATCAGAGG 197
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 134 AAGAAGACCTGCAGGTGTGTCGGCTCTTTTAACTAGAGGCGCTATTAT-----AA 184
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 198 CTTAATTTCTTGACACGCCAGAGAGAGAGCCCAAGCCGTAGTTACTCAGAGCAGCGGAAA 257
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 185 ATTGAAGCAGCTTTCAAAAGAAACGACGAAACGGGGTGTGCGCCAGTGCAGGAAA 244
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 258 CCATGCCCAAGCTCTCACCTATGCTGCTAACTTGAAGAAATTCCTGCTTACATTTGCTG 317
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 CCATGCGCAGGAGTGCCTTTTCCCTGCAACATCTCGGTATTACCGGAAAGATTTTAT 304
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 318 TCCCCAAACAGTCCCAACTCGAAGAACTGGCAATCCAGCCTATGGAGCATCGATAGT 377
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 305 GCGGTGCGAGCAGCCCGAGGCAAAAAATTTCCCAAGTCGAATGTTCCGCAAGGAATACGT 364
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 378 ATACTGTGACCCCAAGTGACGAGTCCAGAGAAAGGTCACCTCAAGAAATATG----- 429
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 365 CGAGATCATCTGACAGGCGACACGTTTGATGATGTTTATCAAGGCGAGTGCCTGCTG 424
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 430 -CAAGAAACAGAGGATCTTGGTTCATCCCAACAGGAGCGCTGAGTATAGCTGGACA 488
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 CGAGGAAGAAAAGCGCGCATTTATCCACCGTTTGACGATCCTGCGCGTATGCTGGCCA 484
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 489 AGGAACAATTTGCCCTGGAAGTGTGAACAGGTTCCCTTGGTA---GATGCACTGGTGT 545
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 485 GGGTACGGTGGCGGTGCAATCTTAAACGATATCGAAACGGAACCCCATTTATTATTCG 544
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 546 ACCAGTAGGAGGAGGAGGATGTTGCTGGAATAGGCATTTACAAATTAAGGCCCTGAAC 605
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 545 AAGCGTAGGAGGAGGAGGCTGCTTTACAGAGTGGGAACCTATATGAAAGAGCTGTCC 604
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 606 TAGTGTGAAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACTGCTACAGTCTAAACT 665
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 605 GGATACAAAGCTGATCGCGTTGAGCCTAAAGCGCGCGGCTCTTTTGTGAATCAACAA 664
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 666 GAAGAGGAACTGACCCCAATCTTCATCCTCCAGAAACCATAGCAGATGGTGTCAATC 725
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 AAAAGCGAAGTGTGCTATCCTTTGAGAAATTCGATAAATTTGCGATGGAGCGGCTGTG 724
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 726 CAGCATTGGCTTGAATACCTGCGCTATTATTAAGAGACCTTGTGGATGATGCTTCACTGT 785
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 725 GAAATCGCGGAAGAGACATTTAAACGCTTGAACCGTTTGCATGATATTTTGCTCGT 784
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 786 CACCGAAG 793
Db      |||||
QY 785 TCCTGAAG 792
Db      |||||

RESULT 12
US-09-939-964-1
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0088
; CURRENT APPLICATION NUMBER: US/09/939,964
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-939-964-1

Query Match          5.2%; Score 52.6; DB 9; Length 536165;
Best Local Similarity 44.5%; Pred. No. 2e-05;
Matches 263; Conservative 0; Mismatches 319; Indels 9; Gaps 1;
```

```
QY 35 TTGAAAAGCTCATATCAACATTCAGACTCTATCCACCTCACCCAGTGCCTAACAGCT 94
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410024 TCAGGCGGCGGTGAGCGGATCGGAAGACAGCTCTTCGCTACCCCTCTGAGACATCAA 410083
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 CCATTTTGAATCAATAGCAGGCGGCAATCTTTTCTTCAAAATGTGAGCTCTTCCAGAAA 154
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410084 GGTGCGTAACCGAACTTACCGGAACCTCAGCTCAGCCTCAAGCTCGAGCACTATCAGCGCA 410143
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 155 CTGGGCTCTTTAAGATTCAGGTGCGCTTAATGCCATCAGAGGCTTAATCTTGACACGC 214
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410144 CTGCTAGCTTTAAGCTGCGGGTGGCAAAACGCAA-----TTCTTCAACTCAGCC 410194
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 215 CAGAAGAGAAAGCCCAAGCCGTAGTTACTCAGAGCAGCGGAACCATGCCAGCTCTCA 274
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410195 CGTCGATCGGCGACGTGGGGTTATTGCGGATCTACGGGCAATCACGGACGGCTCTTT 410254
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 275 CCTATGCTCTAACTGGAAGAAATTCCTGCTTACATTTGTTTCCCAACAGCTCCCA 334
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410255 CCTACGCCCAAAAGCGTGGCTCTCGCGCCACCATCTGCATCTGCGGATCTTGTTCAG 410314
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 335 ACTGCAAGAAACTGGCAATTCAGCCTATGGAGATCGATAGTACTGTGACCCCAAGTG 394
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410315 AAAACAAGGTTTCCGAGATCCGGAAGCTTGGCGGACAGTTCGGATAGTGGATCGTCAC 410374
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 ACGAGTCCAGAGAAAAGGTCACCTCAAGAAATTTATCAAGAAACAGAACGCTCTTGGTCC 454
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410375 AAGACGATCGGCAAGTCGAAGTCGAGCGGCTCGTCGGGAGGAAGGCTTCAGATGATCC 410434
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 455 ATCCCAACAGGAGGCTGCAGTGTAGCTGGAAGAAATTCCTTCCCTGGAAGTCTGA 514
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410435 CGCCTTTTCATCACCAGCATATCATCGCGCGCCAGCAACCGCTCGGTCTTTGAGATCGTTG 410494
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 515 ACCAGGTTCCCTTGGTAGATGCATGCTGGTGTACAGTAGGAGGAGGAGTGGTGTG 574
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410495 AGGCGATGCCGAGCGTCCGATGCTGCTTCCACTGTGCGGTGGCGCTTGGTGCAG 410554
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 575 GAATAGCATTACAAATTAAGCCCTGAAACCTAGTGTGAAGGTATACGGTGG 625
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 410555 GCCTGAGCAGCGGTGAAGCAGCTGCGGCTCATGCGAGGATCATCGGTG 410605

RESULT 13
US-09-974-300-4809
: Sequence 4809, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: APPLICANT: Clausen, Ib Groth
: TITLE OF INVENTION: Methods for Monitoring Multiple Gene
: FILE REFERENCE: 10085.500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: CURRENT FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: 09/680,598
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/279,526
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4809
: LENGTH: 926
: TYPE: DNA
: ORGANISM: Bacillus clausii
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(926)
: OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4809

Query Match 4.5% Score 46; DB 10; Length 926;
Best Local Similarity 44.2%; Pred. No. 6.6e-05;
Matches 346; Conservative 0; Mismatches 415; Indels 21; Gaps 3;

QY 31 GATGTTGAAAAGCTCATATCAACATTCAAAGACTCTATCCACCTCAGCCCGAGTGCTAAACA 90
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 6 GATATTATTCGCAAAATCAACAAATTAAGATGTGTCGACGATACACCTTTGCAAAAA 65

QY 91 AGCTCATTTTGAATCAATAGCAGCGCGCAATCTTTTCTTCAATGTGAGCTCTCTCCAG 150
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 66 GATCAGGTTTGTCTGAACGATATGATGTCACAGTTTACTTAAACGGGGAAGACTTGCAG 125

QY 151 AAAAAGTGGTCTTTTAAGATTCGAGGTGCGCTTAATGCCATCAGAGGCTTAATTCCTGAC 210
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 126 GTTGTGAGATCTTTTAAATTCGGCGGCTATTACCAATTTCTTCGCTAT----- 177

QY 211 ACGCCAGAGAGAGCCCAAGCCGTAGTTACTCAGCAGCGGGAACCAACATGGCCCAAGCT 270
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 178 -CAAGAAGAACTTGCCTGCGTGTATGCGCAAGCGCTGGAACCAACATGCACAAGGG 236

QY 271 CTCACCTATCTCTAACTGGAAGGAAATTCCTGCTTACATGTTGGTTCCTCCCAACAGCT 330
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 237 GTTGCCTATTCATGCGCGCGCTTTAAAGTCAAAAGGTGTTATTTTCATGCGCGACTACCAG 296

QY 331 CCCAACTGCAAGAACTGGCAATCCCAAGCTATGAGCATCGATAGTATATCTGTGACCCA 390
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 297 CCAAGCAAAAGTCGCACAAAGTGAATTTTGGCAGGAGCTATGTAGACGTACGGTTA 356

QY 391 AGTGACGAGTCCAGAGAAA-----GGTCACTCAAAGAAATTTATGCAAGAAACAGAA 441
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 357 ATTGGCGATACGTTTGATGATTTCTTATGCGGCGCCATTCATATTTGCGAAGAACAGGAA 416

QY 442 GGCATCTTGTCATCCCAACAGGAGCGCTGAGTACGTAGCTGGACAAAGCAAAATTTGCC 501
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 417 ATGACGTTTATCCATCCATTTTAAACAGACAAAGTGTATTCGCGGGAAGCAAGCTCGGA 476

QY 502 CTGAAGTGTCTGA---ACCAGGTTCCCTTGGTATGATGCTGCTGCTGCTGCTGCTGCTGCTG 558
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 477 CTTGAAATATGATGACATCGAGGAAACCCAGATTTTGTGTTTTCATCAATTCGGCGGC 536
```

```
QY 559 GGAGGAATGTTGCTGGAATAGCCATTACAATTAAGCCCTGAAACCTAGTGTGAAGGTA 618
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 537 GGTGGGCTTATTAGTGGGATGGCCACCTATATAAAAAAGCGTTACGCCCAACTACGAAAATG 596

QY 619 TACGCTCTGAGCCCTCGAATGAGATGACTGTCTACCACTGCTAAACTGAAGAGAACTG 678
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 597 ATTGGTTGTGAACCTGCTGTGTCATCATGACGGAATCAATTAAGCAAGGCAACGTT 656

QY 679 ACCCCCAATCTTCATCTCCAGAAACCATAGCAGATGCTGCAAAATCCAGCATTTGGCTTG 738
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 657 GTTGAGTTAGACGAATTCGAAATTTGTGATGGAGCGGCTGTGAAAAAAGTCGGCGAC 716

QY 739 AATACCTGGCTATTATAAGAGACCTTTGTGGATGATGCTTCTCACTGTCACCGCAAGATGAA 798
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 717 AAAACACTGCAAAATTTGCAAGAAGTCTGTAGATGACATTTATTGTTGTGCCCAAGAGGCAAA 776

QY 799 AT 800
  ||
Db 777 AT 778
  ||

RESULT 14
US-10-098-841-197
: Sequence 197, Application US/10098841
: Publication No. US20020197679A1
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Xu, Chongjun
: APPLICANT: Zhou, Ping
: APPLICANT: Ma, Yuning
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhao, Qing A.
: APPLICANT: Ren, Felyan
: APPLICANT: Chen, Rui-hong
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhang, Jie
: APPLICANT: Qian, Xiaohong B.
: APPLICANT: Drmanac, Radoje T.
: TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
: TITLE OF INVENTION: Polypeptides
: FILE REFERENCE: 784CIP2
: CURRENT APPLICATION NUMBER: US/10/098,841
: CURRENT FILING DATE: 2002-03-13
: PRIOR APPLICATION NUMBER: 09/598,042
: PRIOR FILING DATE: 2000-06-20
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 331
: SOFTWARE: PL_FL_genes Version 1.0
: SEQ ID NO 197
: LENGTH: 1379
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (182)..(1171)
US-10-098-841-197

Query Match 3.88; Score 38.8; DB 9; Length 1379;
Best Local Similarity 49.08; Pred. No. 0.027;
Matches 143; Conservative 0; Mismatches 137; Indels 12; Gaps 1;

QY 42 AGCTCATATCAACATTCAGACTCTATCCACCTCACCCAGTGTAAACAGCTCAATTTT 101
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 199 AGAGCATGCCAAGCAGGAGGCCCTTTCACGTGGTCACACCTCTGTGTGGAGAGCTGGGCGCT 258
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:11:01 ; Search time 276.093 Seconds  
(without alignments)  
132.862 Million cell updates/sec

Title: US-09-889-609B-8

Perfect score: 1740

Sequence: 1 MCAQYCISFADVEKAHINQ.....TSLNWNQGAERPAPIQTVSV 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB\_PEP.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB\_PEP.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB\_PEP.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB\_PEP.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB\_PEP.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB\_PEP.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB\_PEP.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB\_PEP.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB\_PEP.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB\_PEP.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB\_PEP.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB\_PEP.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB\_PEP.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB\_PEP.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	357.5	20.5	310	9	US-09-738-626-4591
2	307	17.6	436	9	US-09-738-626-3828
3	201.5	11.6	306	10	US-09-815-242-13657
4	196	11.3	308	9	US-09-738-626-6317
5	177	10.2	325	9	US-09-931-457A-32
6	175.5	10.1	310	10	US-09-815-242-10707
7	168	9.7	323	10	US-09-815-242-10239
8	167	9.6	383	9	US-09-931-457A-65
9	161	9.3	323	10	US-09-815-242-13862
10	160.5	9.2	325	9	US-09-931-457A-31
11	160.5	9.2	386	9	US-09-931-457A-66
12	160	9.2	398	9	US-09-931-457A-62
13	159.5	9.2	313	10	US-09-815-242-12146
14	159.5	9.2	313	10	US-09-815-242-12800
15	157.5	9.1	303	10	US-09-815-242-5437
16	157	9.0	415	9	US-09-931-457A-64
17	156.5	9.0	324	9	US-09-905-290A-4
18	154.5	8.9	324	10	US-09-815-242-11879
19	143.5	8.2	316	10	US-09-815-242-11179

20	141.5	8.1	314	10	US-09-815-242-5737	Sequence 5737, Ap
21	129	7.4	306	10	US-09-815-242-11292	Sequence 11292, A
22	115.5	6.6	317	9	US-09-738-626-5838	Sequence 5838, Ap
23	111.5	6.4	250	10	US-09-815-242-11635	Sequence 11635, A
24	110.5	6.4	853	10	US-09-815-242-11363	Sequence 11363, A
25	110.5	6.4	26926	9	US-09-739-508B-2	Sequence 2, Appli
26	109	6.3	1187	10	US-09-935-291A-4	Sequence 4, Appli
27	105.5	6.1	852	10	US-09-815-242-11535	Sequence 11535, A
28	102	5.9	164	10	US-09-734-017A-28	Sequence 28, Appl
29	94.5	5.4	2382	9	US-10-196-935A-2	Sequence 2, Appli
30	93.5	5.4	540	9	US-09-742-096-22	Sequence 22, Appl
31	93.5	5.4	1786	9	US-09-742-096-3	Sequence 3, Appli
32	92.5	5.3	202	10	US-09-734-017A-24	Sequence 24, Appl
33	90.5	5.2	212	9	US-09-742-096-25	Sequence 25, Appl
34	90.5	5.2	442	9	US-10-117-417-6	Sequence 6, Appli
35	90.5	5.2	442	9	US-10-117-417-16	Sequence 16, Appl
36	90.5	5.2	630	9	US-09-742-096-5	Sequence 5, Appli
37	89.5	5.1	297	9	US-09-941-947A-20	Sequence 20, Appl
38	89.5	5.1	297	10	US-09-934-903-14	Sequence 14, Appl
39	89.5	5.1	297	10	US-09-934-868-72	Sequence 72, Appl
40	88.5	5.1	259	9	US-09-996-634-133	Sequence 133, App
41	88.5	5.1	259	9	US-09-997-181-133	Sequence 133, App
42	88.5	5.1	259	9	US-09-997-182-133	Sequence 133, App
43	88	5.1	947	9	US-10-101-464A-73	Sequence 73, Appli
44	88	5.1	1493	10	US-09-858-754-4	Sequence 4, Appli
45	88	5.1	1642	10	US-09-925-442-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-738-626-4591  
; Sequence 4591, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4591  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4591

Query Match 20.5%; Score 357.5; DB 9; Length 310;

Best Local Similarity 30.3%; Pred. No. 2.5e-25;

Matches 101; Conservative 52; Mismatches 133; Indels 41; Gaps 9;

OY 7 ISFADVEKAHINQDSIHLTVLTSILNUNTAGRNLFKCELFKTSFKIRGALN--- 62

DB 2 LTLDNDVITAQORTAPHVRRTPLEFA---DPIDGTQIWIKAFLQKCGVKFGKAFNROLA 58

OY 63 -AIRGLIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVVPOTAPNCKKIAIOAY 121

Db 59 ASEGLIDPT-----VGIVAASGNAGLANAFASLSVPATVLVPETAPQVKVDRLKQY 113  
QY 122 GASIVYCDPSDESREKVTQIMQETEGILVHPNQEPVAVTAGQGTIALEVLNQVPLVDALV 181  
Db 114 GATVQOIGSEYBAFEAAQTFESETGALFCHAYDQPDIAAGAGVIGLEIVEDLPVDVTIV 173  
QY 182 VPVGGGWMAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVK 241  
Db 174 VAVGGGLYAGIAAVAA--HDIKVAVPEPSKIPTLHNSLIAGQ-----PVDVNVSGIA 225  
QY 242 S-SIG-----LNTWPIIRDLDVDDFTVTEDEIKYATQLVWGRMKLLIEPTAGVA 289  
Db 226 ADSIGARQIGREAFDIATAHPPIGLVDD-----EAIIAARRHLWDNVRIPAETHCAAAA 279  
QY 290 LAAVLSQHQTQVSPVKNVCIVLSGGNVDLTSL 322  
Db 280 LASLTGAYKPADE--KVAVIVCGANTDLTTL 310

## RESULT 2

US-09-738-626-5828

; Sequence 5828, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738, 626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: Patent in ver. 3.0

; SEQ ID NO 5828

; LENGTH: 436

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-5828

Query Match 17.6%; Score 307; DB 9; Length 436;  
Best Local Similarity 28.1%; Pred. No. 2.2e-20;  
Matches 94; Conservative 58; Mismatches 162; Indels 20; Gaps 9;  
QY 7 ISFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66  
Db 21 IRAADLTQARISSVIATPQYCPRLSEETGAEIYLKREDLQDVRSKIRGALNSGAQ 80  
QY 67 LIPDTPPEKPKAVYTHSSNGHQAITYAAKLEGIPAYIVVPTAPNCKKLAIOAYG---A 123  
Db 81 L---TQQRDAGIVAASAGNHAGVAYVCKSLGVQGRIVVPTQPKQRDRIMVHGFEV 137  
QY 124 SIYVCDPSDESREKVTQIMQETEGILVHPNQEPVAVTAGQGTIALEVLNQVPLV----DA 179  
Db 138 LSVTGNNTDEASAAHAEDARTGATLIEPFDARNTVIGQTVAAAILSQLTSMGRSADH 197  
QY 180 LVYPVGGGWMAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADG 239  
Db 198 VWYPVGGGGLLAGVSYMADMAPRTAIVGIEPAGAA--MQAALHNGGPITLETVPFDVG 256

QY 240 VK--SSIGLNTWPIIRDLDVDDFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQ 296  
Db 257 AAVKRVGDLNYYTIVENQGRVHMMSATEGAVCTEMLDLYQNEGIIAEPPAGALSIALKEM 316  
QY 297 HFQTVSPVKNVCIVLSGGNVDLTSLNWNVQGAER 330  
Db 317 SF---APGSSVWVCII--SGGNNDV--LRYAEIAER 344

## RESULT 3

US-09-815-242-13657

; Sequence 13657, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13657

; LENGTH: 306

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13657

Query Match 11.6%; Score 201.5; DB 10; Length 306;  
Best Local Similarity 25.3%; Pred. No. 1e-10;  
Matches 74; Conservative 53; Mismatches 124; Indels 41; Gaps 9;  
QY 18 NTODSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRGLIPDTPPEKPK 77  
Db 6 NITELIGQTPVKNLVNIVPEGAADVYKLEAFNPGSSVKDRIALSMIEKAEQDILKPGS 65  
QY 78 AVYTHSSNGHQAITYAAKLEGIPAYIVVPTAPNCKKLAIOAYGASIVYCDPSDESREK 137  
Db 66 TIVEATSGNTGTLGSWVGAAKGKVVIVMPETMSVERRKIIQAYGAELVLTGSEGMKA 125  
QY 138 V--TQIMQETEGILVHPNQEP-----VIAQGTIALEVLNQVPLVDALVVP 183  
Db 126 IAKAQEIAAERDGLPLQFDNPNANPEVHERTTGAELIAAFKDKGL-----DAFVAG 176  
QY 184 VGGGWMAGIATITKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADG-VKS 242  
Db 177 VGTGTISGVSHALKSENSNIQVFAVE---ADE--SAILSGE-KPGPHKIQGISAGFIPD 230  
QY 243 SIGLNTWPIIRDLDVDDFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVL 294  
Db 231 TLDTRAY-----DGIVRTSDDALALGREIGGEGFLV-----GISSAAAI 271

## RESULT 4

US-09-738-626-6317  
Sequence 6317, Application US/09738626  
Publication No. US20020197605A1

## GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

PRIOR FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: Patent in ver. 3.0

SEQ ID NO 6317

LENGTH: 308

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-738-626-6317

Query Match 11.3%; Score 196; DB 9; Length 308;

Best Local Similarity 28.4%; Pred. No. 3.4e-10; Indels 58; Gaps 12;

Matches 80; Conservative 35; Mismatches 109; Indels 58; Gaps 12;

QY 18 NIQDSIHLTPVLTSSILNOIAG--RNLFKCELFKFGKIRGALNAINRGLIPDTPEE---74

Db 4 NITETIGHTPLVNLKLTGLDNLVLFKESFNPNANSVKDRIGL-----AIVDAEKSGL 59

QY 75 KP-KAVVTHSSNGHQALTYAAKLEGIPAYIVVPQAPNCKKLAIOAGSIVYC-----128

Db 60 KPGGTIVATSGNTGIGLAFIAAAGYRLIICMPASMSLERRTILRAFGAELVLT 119

QY 129 --DPDESREKVTQRTMQETEGTLVHPNOEPVIAQG-----GTIALEVLNOVP 175

Db 120 MQAKDKADEIVAER-----ENAVLARQFENEANPRVHRDITAKEILEDITD 165

QY 176 -LVDALVVPVGGGVMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLRHPE 234

Db 166 GNVDIEFVASEFTGTTGVTGVQVLKKNADQVTVTPE-----ASPL---LTAGKAGPH 216

QY 235 TIADGVKSSIGLNTWPII--RLVDDVFTVT--EDEIKYATQL 273

Db 217 KI-----QGIGANFIPEVLDRKVLDDVLTVSNEDAIATFSRKL 253

## RESULT 5

US-09-931-457A-32

Sequence 32, Application US/09931457A

Patent No. US20020157132A1

## GENERAL INFORMATION:

APPLICANT: Falco, S. Carl

APPLICANT: Allen, Stephen M.

TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

FILE REFERENCE: BB1116 US CIP

CURRENT APPLICATION NUMBER: US/09/931,457A

CURRENT FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: 09/424,976

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: 60/065,385

PRIOR FILING DATE: 1997-11-12

PRIOR APPLICATION NUMBER: 60/049,406

PRIOR FILING DATE: 1997-06-12

NUMBER OF SEQ ID NOS: 72

SOFTWARE: Microsoft Office 97

SEQ ID NO 32

LENGTH: 325

TYPE: PRT

ORGANISM: Citrullus lanatus

US-09-931-457A-32

Query Match 10.2%; Score 177; DB 9; Length 325;

Best Local Similarity 23.9%; Pred. No. 2.2e-08;

Matches 74; Conservative 53; Mismatches 112; Indels 70; Gaps 12;

QY 18 NIQDSIHLTPVLTSSILNOIAG--RNLFKCELFKFGKIRGALNAINRGLIP 69

Db 11 DVTELTGNTPLV---YLNRVDDCCVARVAAKLEMEPCSSVKDRIGYSMISDAENKGLI- 66

QY 70 DTPEEKPAVVTTHSSNGHQALTYAAKLEGIPAYIVVPQAPNCKKLAIOAGSIVYC 129

Db 67 -TPGE--SVLIEPTSGNTGIGLAFIAAAGYRLIICMPASMSLERRTILRAFGAELVLT 123

QY 130 P-----SDSREK-----VTQRTMQETEGTLVHPNOEPVIAQGOTTI 166

Db 124 PARGMKGAVQKAEIEIKAKTPNSYILOQFENPANPKIHYETTG-----PEIWRGSG-- 173

QY 167 ALEVLNQVPLVDALVVPVGGGVMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGEL 226

Db 174 -----GKIDALVSGITGGVITGAGYKLNQNPNIKLYGVEP-----VSAILSGG 219

QY 227 TPNLHPPETIADG-VKSSIGLNTWPII--RLVDDVFTVT--EDEIKYATQLVWGRMKLLIETP 285

Db 220 KPGPKIQIGIGAGFIPGVLDVN-----LLDEVIQVSSSESIETAKLLAKKGLLVGIS 272

QY 286 AGVALAAVL 294

Db 273 SGAAAAAAAI 281

## RESULT 6

US-09-815-242-10707

Sequence 10707, Application US/09815242

Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10707

; LENGTH: 310

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-10707

Query Match

10.1%; Score 175.5; DB 10; Length 310;

Best Local Similarity 24.7%; Pred. No. 2.9e-08;

Matches 70; Conservative 55; Mismatches 139; Indels 19; Gaps 8;

Qy 18 NIQDSIHLTPVLSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAIRGLIPDTPEPKP 77

Db 8 SVTELIGSPWVKLRMPVPGAADVFKLEFFNPGSVKDRIALSMIQAAEADGRLLPGQ 67

Qy 78 AVYTHSSNGHGOALTYAAKLEGIPAYIVVPOTPAPNCKKLAIOAGSIVYCDSDSREK 137

Db 68 TIEPTSGNTGIGLAWGAALGYPVKIVMPDTSIERKLMQAYGADLLTPGAEGMKA 127

Qy 138 VTO-RIMOETEGI---LVHPNQEPVIAQO--GTIALEVLNQVPLVDALVVPVGGGMVA 191

Db 128 IAKATALAEHGYFMPLOFNPNANPMLHEOKTGEIVDFGKRL-DAFVSGVGTGTVT 186

Qy 192 GIAITIKALPSPVKKVAAEPSNADDCVQSKLKGELTPNLHPPTIADGVKSSIGLNTWPI 251

Db 187 GVGHKRIFFDIEIIVAVEPTES-----PVLEGG-EPGPHKIQIGAGFVPEV-LDT--- 236

Qy 252 IRLVDVDFVTVEDEIKYATOLVWGRMKLLIEPTAGVALAAVL 294

Db 237 --TVYQKVAVSSDEDALETARLMGPKGEGILVGISAGAAIKAAI 277

RESULT 7

US-09-815-242-10239

; Sequence 10239, Application US/09815242

; Patent No. US20020061369A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10239

; LENGTH: 323

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-815-242-10239

Query Match

9.7%; Score 168; DB 10; Length 323;

Best Local Similarity 25.3%; Pred. No. 1.5e-07;

Matches 78; Conservative 53; Mismatches 133; Indels 44; Gaps 12;

Qy 26 TPVLTSSILNQIAGRNLFKFCFELFQKTSFKIRGALNAI-----RGLIPDTPEPKA-V 79

Db 15 TPLVR---LNRIGNGRILAKVESRNPFSVVKCRIGANMIWDAEKRGVL-----KPGVEL 65

Qy 80 VTHSSNGHGOALTYAAKLEGIPAYIVVPOTPAPNCKKLAIOAGSIVYCDSDSREKVT 139

Db 66 VETSGNTGALAYVAAARGYKLTLPETMSIERKLLKALGANLVLTGA-----KGM 120

Qy 140 QRIMQETEGILVHPNQEPVIAQO-----GTIALEVLNQVP-LVDALVVPVGGG 187

Db 121 KGAIQRAEEI-VASNPKEKYLLOQFSNPANPEIHEKTGTGEIWEDEDGQVDVFIAGVGTG 179

Qy 188 GMVAGIATITKALKPSVKV--YAAEPSNADDCVQSKLKGELTPNLHPPTIADG-VKSSI 244

Db 180 GTLTGVSRYIKGTGKRTDLISVAVEPTDSPVIAQALAGEEIKGPHKIQIGAGFIPANL 239

Qy 245 GLNTWPIIRDLVDDVFTVEDEIKYATOLVWGRMKLLIEPTAGVALAAVLSQHFQTVSPE 304

Db 240 DLK-----LVDKVICITNEEAISTARRLMEEGILAGISSGAATAALKLO-EDESFT 291

Qy 305 VKNVCIVL 312

Db 292 NKNIVVIL 299

RESULT 8

US-09-931-457A-65

; Sequence 65, Application US/09931457A

; Patent No. US20020157132A1

; GENERAL INFORMATION:

; APPLICANT: Falco, S. Carl

; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

; FILE REFERENCE: BB1116 US CIP

; CURRENT APPLICATION NUMBER: US/09/931,457A

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: 09/424,976

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/065,385

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/049,406

; PRIOR FILING DATE: 1997-06-12

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 65

; LENGTH: 383

; TYPE: PRT

; ORGANISM: Spinacia oleracea

US-09-931-457A-65

Query Match

9.6%; Score 167; DB 9; Length 383;

Best Local Similarity 23.7%; Pred. No. 2.4e-07;

Matches 77; Conservative 58; Mismatches 132; Indels 58; Gaps 14;

Qy 6 CTSFADVERAHINIQDSIHHTPVLTSILN-----QIAGR-----NLFF 44

Db 38 CNNFKVSSPITCK-AVSLSPSTIEGLNIAEDVSQLIGKTPMYLNNVSKGSVANIAA 96

Qy 45 KCFLFOKTSFKIRGALNAIRGLIPDTPEK-----PKAVVTHSSNGHGOALTYAAKLE 98

Db 97 KLESMEPCCSVKDRIGYSMI-----DDAEQKGVITPGKTTLVETPSGNTGIGLAFTAAAR 151

Qy 99 GIPAYIVVQTPAPNCKKLAIOAGSIVYCDSDSREK---EKVQTRIMQETEGILVHPNQ 155

Db 152 GYKITTLPASMSMERRVILKAFGAELVLTDPAKGMKGAVERAEELKTKTPDPSYMLQQFD 211

Qy 156 EPA-VIAGOGGTALEVLNQVP-LVDALVVPVGGGMVAGIATITKALKPSVKVYAAEPSN 213

Db 212 NPANPKIHVETTPGEIWEDETKGVDIFVAGICTGGTISGVGRYLKERNFCVQVIGIEPT- 270



QY 214 ADCVOSKUKGELTPTNHLHPETIADG-VKSSIGLNTWPIIRDLVDDVFTVTDE-IKYAT 271  
Db 271 -----ESNLISGCKPPIHQIGLGFVPSNLDLG-----VMDEVIESSEEAVERMAK 318  
QY 272 QLVWGRMK--LLIEPTAGVALAVAL 294  
Db 319 QLA---MKEGLLVGISSGAAAAAVAL 340

## RESULT 9

US-09-815-242-13862  
; Sequence 13862, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: fastseq for Windows Version 4.0  
; SEQ ID NO 13862  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(323)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-13862

Query Match 9.3%; Score 161; DB 10; Length 323;

Best Local Similarity 24.1%; Pred. No. 6.9e-07;  
Matches 74; Conservative 55; Mismatches 136; Indels 42; Gaps 11;

QY 26 TPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAI-----RGLIPDTPEEKPKA-V 79  
Db 15 TPLVR---LNRIGNGRILAKVESRNFVSFKRIGANMTWDAEKRGL-----KPGVEL 65  
QY 80 VTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGASTIVCDPSDESREKVT 139  
Db 66 VEPTSGNTGIALAYAAARGYKLTMTPEMTSTERRKLLKALGANLVLTGA-----KGM 120  
QY 140 QRTMQETGILVHPNQEPAVI-----ACQGTIALEVLNQVP-LVDALVVPVGGGG 188  
Db 121 KGAIOKAEETVADPOKYLLOOFXSPANPEIHEKTTGPEIWDTDGQVDVFTSGVGTGG 180  
QY 189 MVAGIATIKALKPSVKV--YAAEPSNADDCYOSKLGKELTPNHLHPETIADG-VKSSIG 245  
Db 181 TLTGVTTRYIKGTGKTDLTITVAVEPTSPVIAQALAGEEIKPGPHIQIGIGAFIPGNLD 240

QY 246 LNTWPIIRDLVDDVFTVTDEIKYATQLVWGRMKLLIEPTAGVALAVALSQHFOQTVSPEV 305  
Db 241 LK-----LIDKVVGITNEEAISTARRLMEEGILAGISSGAVALAKLQ-EDSESFTN 292  
QY 306 KNCVIVL 312  
Db 293 KNIVWIL 299

## RESULT 10

US-09-931-457A-31  
; Sequence 31, Application US/09931457A  
; Patent No. US20020157132A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl.  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes  
; FILE REFERENCE: BB1116 US CIP  
; CURRENT APPLICATION NUMBER: US/09/931,457A  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: 09/424,976  
; PRIOR FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: 60/065,385  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/049,406  
; PRIOR FILING DATE: 1997-06-12  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 31  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-931-457A-31

Query Match 9.2%; Score 160.5; DB 9; Length 325;

Best Local Similarity 25.5%; Pred. No. 7.8e-07;  
Matches 83; Conservative 54; Mismatches 131; Indels 57; Gaps 16;

QY 12 VEKAHI--NIQDSIHLPVLTSSILNOIAG---RNLFKCELFQKTSFKIRGALNAI-- 64  
Db 3 VERSGIADKDVTELLIGTKPLV---YLNKLADGCVARVAAKLELMPECCSVKDRIGYSMIAD 59  
QY 65 ---RGLIPDTPEEKRAVVTTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIOAY 121  
Db 60 AEEKGLI--TPGK--SVLIEPTSGNTGIGLAFMAAARGYKLIITMPASMSLERRILLAF 115  
QY 122 GASIVYCDPSDESREKVTQRMQETECILVH-PNQEPAVIAGQ-----GTIALE 169  
Db 116 GAELVLTDPF-----KMGAVQKAEIILAKTN---AYILQOFENPANKVHYETTGPE 167  
QY 170 V-LNQVPLVDALVVPVGGGVMAGIATITIKALKPSVKVYAAEPSNADDCYOSKLGKELTP 228  
Db 168 INKSGDKIDAFVSGIGTGTGTGAGKYLKEQNPKNILIGVEP-----VESPVLSGGKP 221  
QY 229 NLHPPETIADG-VKSSIGLNTWPIIRDLVDDVFTVTDEIKYATQLVWGRMKLLIEPTAG 287  
Db 222 GPHKIQIGAGFPGVLEVN-----LLDEVVQISSDEAIEATAKIALKEGLFVGISGG 274  
QY 288 VALAVALSQHFOQTVS-PEVKNVCIV 311  
Db 275 AAAAAA-----FQIAKRPENAGKLIV 295

## RESULT 11

US-09-931-457A-66  
; Sequence 66, Application US/09931457A  
; Patent No. US20020157132A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl.  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes  
; FILE REFERENCE: BB1116 US CIP  
; CURRENT APPLICATION NUMBER: US/09/931,457A

```

Qy 23 IHLTPVLVTSILNQIAGRNLFKFKELFOKTSFKIRGALNAI-----RGLPDTPEPK 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 VVLNNVVKGSYANVA-----KLEIMPPCCSVKDRIGYSMINDAEQGLI--TPGK--S 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 78 ANVTHSSGNHGQALTYAAKLEGIPAYIVVPQTPANCKKLAIOAGSIVVYCDPDES 135
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 VIVEATSGMTGTGLAFIAASKGYKLIITMPPSSMSMERRVLLRAFGAELVITDAKGMKA 204
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 136 -EKVQRIOMQETEGILVHPNQEPA-VIAGQCTIALEVL-NQVPLVDALVVPVGGGVMVAG 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 LDKATEILNKTPNSYMLQOFDPNPANPQVHYETGTPEIWEDSKGKVDFIGICTGTISC 264
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 193 IAITTKALKSPVYVAAEPSNADCCYOSKLGELTPMLHPETIADGVKSSIGLNTWPII 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 AGRFLKEKNPGTKVIGIEPS-----ESNILSGKPGPKHQIGIAGF---VPRN---LD 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 253 ROLVDDVFTVTEDE-IKATQLVWGRMKLLIETPTAGVALAAVL 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 SDILVEVIEISSDEAVETAKOLA-VQEGLLGVISSGAAAAAAI 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-815-242-12146
; Sequence 12146, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12146
; LENGTH: 313
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12146

Query Match 9.28; Score 159.5; DB 10; Length 313;
Best Local Similarity 23.6%; Pred. No. 9.1e-07;
Matches 70; Conservative 52; Mismatches 151; Indels 23; Gaps 8

Qy 9 FADVCKEAHNTIODSTHLTPVLTSITLNCIAGRNLFKFKELFOKTSFKIRGALNAIRGLI 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 FLMAQKPVDNITLIIGGTPVVKLRNVDDNADVVYKLEYQNPGSVKDRIALAMIEKAE 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 PTPPEEKPAKVYTHSSGNHGQALTYAAKLEGIPAYIVVPQTPANCKKLAIOAGSIVV 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 RGCKIKPGDTIIVPTSGTGTGLAFVCAAKGYKAVFTMPETMSQERRNLLKAYCAELVLT 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

```

RESULT 14
US-09-815-242-12800
; Sequence 12800, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes

```

	Query Match	9.2%; Score 159.5; DB 10; Length 313;
	Best Local Similarity	23.6%; Pred. No. 9.le-07;
	Matches	70; Conservative 52; Mismatches 151; Indels 23; Gaps 8;
QY	9 FADVERAHINIQDSIHLTPVLTSSILNQIAGRNLFFKCELLFQKTGSEFKIRGALNAIRGLI	68
DB	2 FLMAQRPVDNITQIGCTPPVKLVNVDNDAADVVKLEYQNPGSVKDRIALAMIEKAE	61
QY	69 PDTPEEKP KAVHTSSNGHQALTYAAKLGP IAYIVVPOTAP NCKKLAIQA YGASIVYC	128
DB	62 REGKIKPGDIVEPTSGNTIGLAFVCAAKGYKAVFTMPETMSOERNLLKAYGAELVL T	121
QY	129 DPDESREKVTO-RIMOTEGILVHPNQ-----EPAVIAGGGTIALELVNQV--PLVDAL	180
DB	122 PGSEAMKA IKKAKELKEEHGF-EPOQFN PANPEV--HELTGPELLOQFEKTTIDAF	178
QY	181 VWPVGGMGVAGTAITTKALKPSVKVYAEPFSNADD CYQSKLGELTPNLHPPETIADGV	240
DB	179 LAGVGGTGTLSSGVKKLVKKEYPINVAIPE-----ASPVLSSGGP PPHKLOGLAGGF	232

[illegible]

Query Match	9.14;	Score 157.5;	DB 10;	Length 303;
Best Local Similarity	23.7%;	Pred. No. 1.3e-06;		
Matches	68;	Conservative 51;	Mismatches 145;	Indels 23; Gaps 8
Qy	18	NIQDSIHLPVLTSSILNQIAGRNLFKFCELFQKTSFKIRGALNAIRGLIPDTPPEERPK	77	
Db	2	NITQIIIGTPVKLVNRVDDNAADVVKLEYQNPGPSVKDRIALAMIEKAEREGIKPGD	61	
Qy	78	AVYTHSSNGHQALTYAAKLEGIPAYIVVPOTAPNCKKLATQAYGASTIVYCDPDSERK	137	
Db	62	TIVEPTSGNTGIGLAFCAAKGYKAVFTMPETMSOERNLLIKYAGLEVLTPGSEAMKA	121	
Qy	138	VTQ-RIMQETEGILVHPNQ-----EPAVIAQGGITALEVLNQV--PLVDALVVPVGGGM	189	
Db	122	IKKAKELKEEHGYF-EQQQFENPANPEV--HELTTPGPELLOQFEGKTTDAFLAGVCTGTT	178	
Qy	190	VAGTATITKALKPSVKVYAAEPSNADDCYQSKLAGELTPNLHPPTIADGVKSSITGLNT	249	
Db	179	LSGVGKVLKREYNPIEIVAEIPE-----ASPVLSGGEPGPKLQGLAGFIPGTT-LNT--	230	
Qy	250	PIIRDLVDDVFTVTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLQSQ	296	
Db	231	----EYDYSIIKVGNDTAMEMSRRAKEEGELLAGISSGAAITYAAIOK	273	

Search completed: June 24, 2003, 06:16:49  
Job time : 277.093 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 06:07:29 ; Search time 26.9602 Seconds  
(without alignments)  
1208.801 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISPADVEKAHINIQ.....TSLNMGQAERPAPQTQVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	791	45.5	346	2 T04211	probable threonine
2	587	33.7	326	2 S38061	threonine ammonia
3	560	32.2	323	2 T41297	threonine ammonia
4	544	31.3	320	2 E83310	L-serine ammonia-1
5	535	30.7	346	2 F89921	threonine ammonia
6	530	30.5	325	2 T36434	threonine ammonia
7	522.5	30.0	401	2 D72386	threonine ammonia
8	520.5	29.9	324	2 A97687	probable serine/th
9	520.5	29.9	324	2 AF2312	threonine dehydrat
10	517.5	29.7	323	2 F95390	probable threonine
11	507.5	29.2	317	2 T24337	threonine ammonia
12	499	28.7	405	2 D90166	hypothetical prote
13	498.5	28.6	325	2 G87633	threonine dehydrat
14	492.5	28.3	495	2 H84359	threonine dehydrat
15	491	28.2	329	2 AB0897	threonine ammonia
16	476	27.4	329	1 DWECDT	threonine ammonia
17	476	27.4	329	2 D85973	threonine ammonia
18	476	27.4	329	2 E91128	threonine ammonia
19	466.5	26.8	320	2 AG0496	threonine ammonia
20	466.5	26.8	343	2 D97778	threonine ammonia
21	463.5	26.6	504	2 F83603	threonine dehydrat
22	446.5	25.7	403	2 D81355	threonine ammonia
23	444.5	25.5	333	2 H71703	probable threonine
24	441.5	25.4	568	2 E75502	threonine ammonia
25	439	25.2	400	2 AB7700	threonine dehydrat
26	426.5	24.5	336	2 AC3142	threonine dehydrat
27	423	24.3	512	2 T23166	hypothetical prote
28	414	23.8	334	2 B95894	probable threonine
29	413	23.7	600	2 T39516	threonine ammonia

30	409.5	23.5	508	2 S77559	threonine ammonia
31	406.5	23.4	515	2 B83479	threonine dehydrat
32	403.5	23.2	304	2 A98146	probable threonine
33	399	22.9	349	2 C72630	threonine ammonia
34	398	22.9	514	2 AG0474	threonine ammonia
35	395.5	22.7	576	1 DWBYT	threonine ammonia
36	391	22.5	510	2 E82374	threonine ammonia
37	384	22.1	514	2 AB0924	threonine ammonia
38	382	22.0	514	2 B91217	threonine ammonia
39	380	21.8	514	1 DWEETT	threonine ammonia
40	378.5	21.8	515	2 C86063	threonine ammonia
41	378	21.7	514	1 DWEETS	threonine ammonia
42	374	21.5	422	2 A69644	threonine ammonia
43	371.5	21.4	43	2 A38628	threonine ammonia
44	371	21.3	592	2 T51712	threonine ammonia
45	368	21.1	257	2 AF1839	threonine dehydrat

ALIGNMENTS

RESULT 1

T04211

probable threonine ammonia-lyase (EC 4.3.1.19) T5C23.70 [similarity] - Arabidopsis th  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 21-Jun-2002

C:Accession: T04211

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Dueterhoeft, A.; Bancro

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15261

A:Accession: T04211

A:Molecule type: DNA

A:Residues: 1-346 <BEV>

A:Cross-references: EMBL:AL049500

A:Experimental source: cultivar Columbia; BAC clone T5C23

C:Genetics:

A:Map position: 4

A:Introns: 55/2; 99/2; 176/2; 263/2

A:Note: T5C23.70

C:Superfamily: threonine dehydratase

C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

Query Match 45.5%; Score 791; DB 2; Length 346;

Best Local Similarity 46.9%; Pred. No. 2.3e-54;

Matches 159; Conservative 68; Mismatches 88; Indels 24; Gaps 6;

Qy 4 QYCISFADVEKAHINIQDSIHLTPVLTSILNIOIAGRNLFKCELFQ----- 50

Db 7 KYAADILLSKEAHDRIKPIHRTPLVLTSESLNSTSGRSLFFKCECLQKGIERSFYFSV 66

Qy 51 --KTGSFKIRGALNIRGLIPDTPPEKPAVVTSSNGHGOALTYAAKLGIPAYIVVPO 108

Db 67 CGKSGAFKFGACNVLSL---DAEQAAKGVVTHSSGNHAAALSLAAKIOGIPAYIVVPR 123

Qy 109 TAPNCKKLAIQAYCASIVYCDPDESREKVTQIMQETEGILVHPNDEPAVIAQOGTIAL 168

Db 124 GAPCKVDNVIRYGGKVINSEATMSREETASKVLQETGSLVHPYNDGRIISQGTIAL 183

Qy 169 EVLNQVPLVDALVVPVGGGGMWAGIATIKALPKSVKVAAPSNADDCYQSKLKGELTP 228

Db 184 ELLEQIQEIDAIVVPISSGGLISGVALAAKSIRISIRIAAEKPGADDAQSKVAGKII- 242

Qy 229 NLHPPETIADGVKSSIGLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLIEPTAGV 288

Db 243 TLPVTNTIADGLRASLDLTPVVPVRLVDVDDVLTLEECEIIEAMKMCYIILKVSVEPSGAI 302

Qy 289 ALAAVLSQHFQTVSP---EVKNVCIVLSGGNVDLTSLNW 324

Db 303 GLAAVLSNFRN-NPSCRCKNIGIVLSGGNVDLTSLW 339

RESULT 2

S38061

threonine ammonia-lyase (EC 4.3.1.19) YK1218c [similarity] - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D326  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jun-2002.  
C:Accession: S38061; S44320  
R:Alexandraki, D.; Tzermia, M.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38054  
A:Accession: S38061  
A:Molecule type: DNA  
A:Residues: 1-326 <ALE>  
A:Cross-references: EMBL:Z28218; NID:g486390; PIDN:CAA82063.1; PID:g486391; MIPS:YK1218c  
A:Experimental source: strain S288C  
R:Tzermia, M.; Horaitis, O.; Alexandraki, D.  
Yeast 10, 663-679, 1994  
A:Title: The complete sequencing of a 24.6 kb segment of yeast chromosome XI identified e dehydratases, membrane transporters, hydantoinases and the phospholipase A(2)-activator  
A:Reference number: S44319; MUID:95028164; PMID:7941750  
A:Accession: S44320  
A:Molecule type: DNA  
A:Residues: 1-326 <TZE>  
A:Cross-references: EMBL:X75951; NID:g473130; PIDN:CAA53555.1; PID:g473132  
C:Genetics:  
A:Gene: SGD:SR1  
A:Cross-references: SGD:S0001701  
A:Map position: 11L  
C:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

Query Match 33.7%; Score 587; DB 2; Length 326;  
Best Local Similarity 40.9%; Pred. No. 2e-38;  
Matches 128; Conservative 66; Mismatches 111; Indels 8; Gaps 4;

Qy 8 SFADVEKAHNIQDSIHLTPVLTSIIINQTAGNRLFFKCELFQKTSFKIRGALNAIRGL 67  
Db 5 TYGDVLDAENRIKEYNKNTPVLTSRMLDRLGAIQIVFKGENFORVCAKFRGAMNAVSKL 64  
Qy 68 IPDTPEEKRAVYTHSSGNHGQALTYAAKLEGIPAYIVVPQAPNCKKLAIQAYGASIVY 127  
Db 65 ---SDEKRSKGVIAFSSGNHAQIALSAKLLNVPATIVMPEDAPALKVAATACGYAHIR 121  
Qy 128 CDPDSREKVTORIMQETEGILVHPNQEPAVTAGGCTTIALEVLNQVPLVDALVVPVGGG 187  
Db 122 YNRYTDEQIGQLAAEHGAFALTPPYDHPDVJTAGGTSAKELLEEVGOLDALFVPLGGG 181  
Qy 188 GWVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246  
Db 182 GLLSGSAARSLSPCKKIFGVPEAGNDCQQQFSRGSIV-HINTPKTIADGAQTOHLGE 240  
Qy 247 NTWPIIRDLVDVFTVTEDEIKYATOLVGMKLLIEPTAGVALAAAVLSQHFQTVSPEVK 306  
Db 241 YTEAIIRENVDILTYSQDELVKCMHFIAERMKVVVEPTACILGFAGALLKKELVG---K 297  
Qy 307 NVCIVLSGGNVDL 319  
Db 298 KVGIIISGGNVDM 310

RESULT 3  
T41297  
threonine ammonia-lyase (EC 4.3.1.19) SPCC320.14 [similarity] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jun-2002  
C:Accession: T41297; T41325  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Wedler, E.; Wedler, E.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: 421986  
A:Accession: T41297  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-323 <WOO>  
A:Cross-references: EMBL:AL022245; PIDN:CAA18316.1; GSPDB:GNO0068; SPDB:SPCC320.14  
A:Experimental source: strain 972H-; cosmid c320

Db 6 TYDDVIAAARIAIAGHANRTVPMSSRTLDEELGAEVFFKCNELQRMGMFAKFRGFALNSRF 65  
QY 68 IPDTPPEKPAVYTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIV 127  
Db 66 ---SAGORAGVAFSSGNHQAIALSARLLGIPATIVMPADAPAVKIETRYGGQVVL 122  
QY 128 CDPDSRESKVTQRIQMOETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVPVGG 187  
Db 123 YDRTYEDREQIGDLAQRHGLTLPYPYDPVLAGOGTAAKELFEVGLDAFFAPLGG 182  
QY 188 GWAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGLTLPNHPPETIADGVKSS-IGL 246  
Db 183 GLUSGALAIIRALAPACRIYGVPEAGNDGQSRSGAIV-HDITPTTADGATQHLGN 241  
QY 247 NTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFOTVSPEVK 306  
Db 242 LTFPLIQRNVDDILTASDAELVDGMRFLAARMKLLVEPTGCLGLAAARQKDEL---RGK 298  
QY 307 NVCIVLSGGNVDL 319  
Db 299 RVGILLSGGNVDL 311

RESULT 5  
F89921  
threonine ammonia-lyase (EC 4.3.1.19) [similarity] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 02-Aug-2002  
C:Accession: F89921  
R:Kuroda, M.; Ohca, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F89921  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-346 <KUR>  
A:Cross-references: GB:BA000018; PID:g13701236; PIDN:BABA2531.1; GSPDB:GNO0149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA1271  
A:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase

Query Match 30.7%; Score 535; DB 2; Length 346;  
Best Local Similarity 37.8%; Pred. No. 2.6e-34;  
Matches 122; Conservative 64; Mismatches 131; Indels 6; Gaps 4;

QY 5 YCISFADVEKAHINIQDSIHLTPVLTSSILNQ-IAGRNLFKCELFQKTSFKIRGALNA 63  
Db 12 HIVSLGDIEAKASIKPFRTPLIKSWLSQNTKGNVYKLENQFTGSGFRGASNK 71  
QY 64 IRLGIPDTPPEKPAVYTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYGA 123  
Db 72 INHL---SDQKAKGIIGASAGNHAQGVALTAKLLGIDATIVMPETAPIAKONATKGYGA 128  
QY 124 SIYCDPSDESREKVTQRIQMOETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVP 183  
Db 129 KVLKGNFNETRLYMEELAKENGMTIVHPYDKFYMAGOGTIGLEILDDIMNVNVIVP 188  
QY 184 VGGGMVAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGLTLPNHPPETIADGVKSS 243  
Db 189 VGGGLIAGIATALKSFNSIHIIGVQAEVNHGMAESFYKRALTEH-REDSTIADGCDVK 247  
QY 244 I-GLNTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFOTVS 302  
Db 248 VPCEKTYEVVKHLVDFEILVSEIEHAMQDLQRAKIITEGAGALPTAAILSGKIDKKW 307  
QY 303 PEVKNVCIVLSGGNVDLTSLNWV 325  
Db 308 LEGKNVALVSGGNVDLTRVSGV 330

RESULT 6  
T36434  
threonine ammonia-lyase (EC 4.3.1.19) SCF43A.11c [similarity] - Streptomyces coelicol  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jun-2002  
C:Accession: T36434  
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21598  
A:Accession: T36434  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-325 <SEE>  
A:Cross-references: EMBL:AL096837; PIDN:CABA48898.1; GSPDB:GNO0070; SCOEDB:SCF43A.11c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCF43A.11c  
A:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase

Query Match 30.5%; Score 530; DB 2; Length 325;  
Best Local Similarity 39.3%; Pred. No. 6e-34;  
Matches 123; Conservative 54; Mismatches 128; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGALNAIRG 66  
Db 8 VTLDVRSAAARKGVNHRTPVLRSTLDALVGAELKCNQORVGAFFRGAJYNAASR 67  
QY 67 LIPDTPPEKPAVYTHSSGNHGOALTAAKLEGIPAYIVVQTPAPNCKKLAIAQYASIV 126  
Db 68 L---TPEQLARGIAAYSSGNHQAQVAALARELGTAVIVMPEDAPPSKRDATAGYGAIV 124  
QY 127 YCDPSDESREKVTQRIQMOETEGILVHPNQEPAVIAGOGTTALEVLNVLVPLDALVVPVGG 186  
Db 125 TYDRYTGDRVAVAAALAAADRLTLIPPYEHPPHYAGOGTAALVELVEETGELDALVAPVG 184  
QY 187 GGMVAGTAITIKALKPSVKVYAAEPSNADDCYQSKLGLTLPNHPPETIADG-VKSSIG 245  
Db 185 GGLIAGSATAVKALHFQMGVIGVEPEAGDDTRSLAARRV-SVPVPTIADGQALDTPG 243  
QY 246 LNTWPIIRDLVDDVFTTEDEIKYATQLVWGRMKLLIETPAGVALAAVLSOHFOTVSPEV 305  
Db 244 ELTFSLNRRLLDGIVLSDDEIRDAIRFAFERLKTVPLEPSGATPLAALLNGRIDALP--- 300

RESULT 7  
D72386  
threonine ammonia-lyase (EC 4.3.1.19) TM0356 [similarity] - Thermotoga maritima (stra  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jun-2002  
C:Accession: D72386  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: D72386  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <ARN>  
A:Cross-references: GB:AE001716; GB:AE000512; NID:g4980853; PIDN:AAD35443.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0356  
A:Superfamily: threonine dehydratase  
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; carbon-oxygen lyase; hydro-lyase; p

```

|::: | |::: | |::: | |::: | |::: | |::: | |::: |
Db      243 NITFPIMAGICGKGIAVSEELRAMVLAFNRLLKVVIEPGAVALAAL---FHGRELES 299

Qy      306 KNVCIVLSGGNVD 318
        : | | | | | | |
Db      300 ETIVASGGNVD 312

RESULT 9
AF2912
theonine dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF2912
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wood-
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCli-
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam-
ster, E.W.
A:title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2912
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43716.l; PID:g17741247; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)

```

A: Map position: circular chromosome  
C: Superfamily: threonine dehydratase

	Query Match	29.9%;	Score 520.5;	DB 2;	Length 324;
	Best Local Similarity	40.6%;	Pred. No. 3.3e-33;		
	Matches 127; Conservative 44;	Mismatches 129;	Indels 13;	Gaps 5;	

  

QY	12	VEKAHINIODSHTLTPVLTSSTLNQTAGRNLPFKCELFKFTGSFKIRGALNAIRGLIPDT	71
Db	7	TEAARERIGNHAVPTPLLTSPFDETAGRKLRFVKAECQLRTSGFRPGWSAVSGLPADV	66
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
QY	72	PEEKRAVVTHSSGNHGQALTAAKLEGIPAYIVVPQTAPNCKKLAIQAYGASIVVCDDPS	131
Db	67	--- RAKGVTAFSGGNHQAQVALAARLHGIPAVIIMPSDAPKTIIDNTRAYGAEEVLYDRA	123
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
QY	132	DESREKVOTIMOETEGILVHPNOEPAVTAGOGTIALEVLNQ-----VPLVDALVVPVG-	186
Db	124	NEDRDIAIGNLSRSRGTLIRPYDEPLVTAGOGTAGCTAEAOGEALGICAAEVL-VPCGG	182
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
QY	187	GMWAGCIATIKALKPSVKVYAAPSNDCCYOSKLUKBELTPNLHPPETIADG-VKSSIG	245

DB 183 GLTGTSISLADAKARNYKVRTAEPERFDVARSAAAGKIERNATTSGSICDAIVTPQPG 242

QY 246 LNTWPTIRLDVDDVFVTVEDETKYAPOLVWGMKLLIEPTAGVALAAVLSSQHFTQVSPV 305

DB 243 NITFPIMAGLCGKIADVSEEAALRAMVLAFNRLKVVIEPGGAVALAAL--FHCKELES 299

QY 306 KNCVICVLSGNVD 318

DB 300 ETVIASVSGNVD 312

RESULT 10

F95390

probable theonine dehydratase [imported] - Sinorhizobium meliloti (strain 1021) maga

C.Species: Sinorhizobium meliloti

C.Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 27-Nov-2001

R.Accession: F95390

R.Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B  
.: Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K  
proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A.Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli  
A.Reference number: A95262; UID:21396509; PMID:11481432







```

||||:||||| ||:|:| | : | ||:| : ||:| | : ||| |
Db 184 GGGLIAGIAIAIKSINPTIKVIGVQAEVHGMAASYAYAGEITAH-RTTGTLAGCDVSR 242
QY 244 IGLNTPPIIRDLDVDDVFTVTEDEIKYATQLVWGRMKLLIJEPTAGVALAAVLS-----Q 296
Db 243 PGNLYEIVRELVDDIVLVSEDEIRNSMLALIQRNKVITEGAGALACAAALLSGKLDSHIQ 302
QY 297 HFQTVSPEVKNVCIVLSGGNVDLTSLNWV 325
Db 303 NRKTVS-----IISGNIDLSRVSQI 323
```

Search completed: June 24, 2003, 06:23:09  
Job time : 28.9602 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:44:58 ; Search time 41.4389 Seconds  
(without alignments)  
1685.614 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYICISFADVEKAHINIQ.....TSLNWNVGOAERPAPYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rviris.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	791	45.5	346	10 Q9T0D1	Q9T0d1 arabidopsis
2	620	35.6	323	16 Q8XPX0	Q8XPX0 ralstonia s
3	546	31.4	388	3 Q8X0J0	Q8X0J0 neurospora
4	544	31.3	320	16 Q910F5	Q910f5 pseudomonas
5	535	30.7	346	16 Q99U50	Q99u50 staphylococ
6	530	30.5	325	16 Q9XAA4	Q9Xaa4 streptomyce
7	522.5	30.0	401	16 Q9WYJ1	Q9wyj1 thermotoga
8	520.5	29.9	324	16 Q8UBW5	Q8ubw5 agrobacteri
9	517.5	29.7	323	16 Q92Y58	Q92y58 rhizobium m
10	507.5	29.2	317	5 Q93968	Q93968 caenorhabdi
11	502.5	28.9	404	17 Q975Y3	Q975y3 sulfolobus
12	501	28.8	402	16 Q8XL77	Q8xl77 clostridium
13	499	28.7	405	17 Q980P1	Q980p1 sulfolobus
14	498.5	28.6	325	16 Q9A3U7	Q9a3u7 caulobacter
15	492.5	28.3	495	17 Q9HNH6	Q9hnh6 halobacteri
16	491	28.2	329	16 Q8Z3K4	Q8z3k4 salmonella

17	487	28.0	404	16 Q8RDT9	Q8rdt9 fusobacteri
18	483.5	27.8	409	16 Q9EWG3	Q9ewg3 streptomyce
19	470	27.0	406	17 Q97CB2	Q97cb2 thermoplasm
20	468	26.9	406	17 Q9HLW2	Q9hlw2 thermoplasm
21	466.5	26.8	320	16 Q8Z9V3	Q8z9v3 yersinia pe
22	466.5	26.8	343	16 Q9ZHZ2	Q9zhz2 rickettsia
23	464	26.7	403	17 Q8ZVF0	Q8zvf0 pyrobaculum
24	463.5	26.6	504	16 Q916G0	Q916g0 pseudomonas
25	453	26.0	469	5 Q9VHF0	Q9vhf0 drosophila
26	446.5	25.7	403	16 Q9PP95	Q9pp95 campylobact
27	444.5	25.5	333	16 Q9ZD93	Q9zd93 rickettsia
28	441.5	25.4	568	16 Q9RW8	Q9rw8 deinococcus
29	440.5	25.3	352	16 Q987A3	Q987a3 rhizobium l
30	439	25.2	400	16 Q9A2D0	Q9a2d0 caulobacter
31	433	24.9	327	16 Q98DQ4	Q98dq4 rhizobium l
32	426.5	24.5	336	16 Q8U6P8	Q8u6p8 agrobacteri
33	425.5	24.5	448	5 Q95XY8	Q95xy8 caenorhabdi
34	423	24.3	512	5 Q21080	Q21080 caenorhabdi
35	419	24.1	334	16 Q985M4	Q985m4 rhizobium l
36	414	23.8	334	16 Q92WC4	Q92wc4 rhizobium m
37	413	23.7	600	3 Q94634	Q94634 schizosacch
38	409.5	23.5	508	16 P73375	P73375 synecocyst
39	406.5	23.4	515	16 Q91418	Q91418 pseudomonas
40	399	22.9	349	17 Q9T8V1	Q9t8v1 aeropyrum p
41	398	22.9	514	16 Q8ZAB4	Q8zab4 yersinia pe
42	391	22.5	510	16 Q9KWI1	Q9kwi1 vibrio chol
43	388.5	22.3	602	10 Q9AUQ1	Q9auq1 oryza sativ
44	388.5	22.3	602	10 Q8W314	Q8w314 oryza sativ
45	384	22.1	514	16 Q8Z378	Q8z378 salmonella

ALIGNMENTS

RESULT 1					
Q9T0D1	ID	Q9T0D1	PRELIMINARY;	PRT;	346 AA.
AC	Q9T0D1				
DT	01-MAY-2000	(TEMBLrel. 13, Created)			
DT	01-MAY-2000	(TEMBLrel. 13, Last sequence update)			
DT	01-MAR-2002	(TEMBLrel. 20, Last annotation update)			
DE	T5C23.70	protein.			
GN	T5C23.70	OR AT4G11640.			
OS	Arabidopsis thaliana	(Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	[1]	SEQUENCE FROM N.A.			
RA	Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,				
RA	Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;				
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.				
RN	[2]	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;				
RA	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.				
RN	[3]	SEQUENCE FROM N.A.			
RA	Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,				
RA	Mewes H.W., Lencic K., Mayer K.F.X.;				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RN	[4]	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;				
RA	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL049500; CAB39935.1; -				
DR	EMBL; AL161532; CAB78207.1; -				
DR	HSSP; P04968; 1TDJ				
DR	InterPro; IPR001926; B6_enzyme_beta.				
DR	InterPro; IPR000634; S/T_dehydrtse.				
DR	Pfam; PF00291; PALP; 1.				
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.				

```
KW Hypothetical protein.
SQ SEQUENCE 346 AA; 36839 MW; 47C37E1CA52B91F0 CRC64;

Query Match 45.5%; Score 791; DB 10; Length 346;
Best Local Similarity 46.9%; Pred. No. 7.9e-55;
Matches 159; Conservative 68; Mismatches 88; Indels 24; Gaps 6;

QY 4 QYCISPADVEKAHINQDSIHLTPVLTSSILNOIAGRNLFKCELFQ----- 50
DB 7 KYAADILSKEAHDRLKPYHRTPVLTSSLSISGRSFFKCECIQKGECSFVSFY 66
QY 51 --KTGFKIRGALNAIRGLIPDPEEKPAVWTHSSNGHQALTYAAKLEGIPAYIVV 108
DB 67 CGKSGAFKRGACNAVLSL---DAEQAAKGVVTHSSNGHAAALSLAAKIQGIPAYIV 123
QY 109 TAPNCKKLAIQAYGASIVYCDPDSREKVTQRMQETEGILVHPNQEPAVTAGQTI 168
DB 124 GAPKQVDNVIRYGGKVIWSEATPMSSREETASKVLQETGSLVHPYNDGRHSQGT 183
QY 169 EVLNQVPLVDALVVPVGGGMAVAGIAITIKALKPSVKVYAAEPSNADDCYQSK 228
DB 184 ELLEQIQEIDAIVVPISSGGLISGVALAKSIKPSIRIITNAEPKGAADAAQSKV 242
QY 229 NLHPPETIADGVKSSIGLNTWPIIRDLVDVFTVTEDEIKYATQLVWGRMKLLI 288
DB 243 TLPVTNTIADGLRASGLDTPVVRDLVDVFTVTEDEICEIEAMKCYEILKVSVE 302
QY 289 ALAAVLSQHFQTVSP---EVKNVCIVLGGNVDLTSLNW 324
DB 303 GLAAVLSNFRN-NPSCRCKNIGIVLGGNVDLGLS-W 339

RESULT 2
Q8XPX0 PRELIMINARY; PRT; 323 AA.
AC Q8XPX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable amino-acid dehydratase protein (EC 4.2.-.-).
DE RSP1516 OR RS04793.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=GMI1000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18667.1; -.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
KW Lyase; Plasmid; Complete proteome.
SQ SEQUENCE 323 AA; 34515 MW; 2175CBF693E04A2D CRC64;

Query Match 35.6%; Score 620; DB 16; Length 323;
Best Local Similarity 41.0%; Pred. No. 2.9e-41;
Matches 133; Conservative 65; Mismatches 118; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINQDSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAIR 66
DB 6 IYDDVVAHARLQGVHFTPLVLTSTANALTGAEILFKCELFQKTSFKIRGALNAIR 65
QY 67 LIPDPEEKPAVWTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGAS 126
```

```
DB 66 F---TPEQRKAGVITPSSNGHQAIALSARLLGMRAVIMPKDAPTIKVEATRGYGEV 122
QY 127 YCDPSDESREKVTQRMQETEGILVHPNQEPAVIACQGTIALEVLNOVPLVDALVVP 186
DB 123 FYDRYTEDREAIGRRILAEQHLTLIPPDYHPHVMAGQTAAKELIETGPDLLLVCL 182
QY 187 GGMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPETIADGVKSS-IG 245
DB 183 GGLLSCCAARALSAPACRIIGVEPAGNDGQSLRKEIV-HIDTPATIADGAQTHLG 241
QY 246 LNTWPIIRLDVDDVFTVTEDEIKYATQLVWGRMKLLIETAGVALAALVLSQHFQTV 305
DB 242 HYTEFAVIRALVDVIATVSADLVDTMRFFAGRMKIVIVPTGCLAAAAALRQR---VEVRG 298
QY 306 KNCVCIVLGGNVDLTSLNWVQAE 329.
DB 299 KRVGVIISGNGVDLQHFARLVQAD 322

RESULT 3
Q8X0J0 PRELIMINARY; PRT; 388 AA.
AC Q8X0J0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Related to threonine dehydratase.
DE B11H24.140.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RC SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670005; CAD21283.1; -.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
SQ SEQUENCE 388 AA; 41505 MW; 02BFB864B95716A6 CRC64;

Query Match 31.4%; Score 546; DB 3; Length 388;
Best Local Similarity 37.5%; Pred. No. 2.9e-35;
Matches 140; Conservative 63; Mismatches 106; Indels 64; Gaps 10;

QY 10 ADVEKAHINQDSIHLTPVLTSSILNOIAGR-----NLFKFC 46
DB 13 ASVLAHSLIKPYIHETPVLTNTFDQLASTPRTPELQGTWEATERPANPKIREWFK 72
QY 47 ELFOKTSFKIRGALNAIRGLIPDPEE---KPAVWTHSSNGHQALTYAAKLEGIPAYI 104
DB 73 ENFORIGAPKAGAFHAEVRLKQTEGLGKGGVVTHSSNGHQAALSAARENGIPAH 132
QY 105 VVPQTPAPNCKKLAIQAYGASIVYCDPDSREKVTQRMQETEGILVHPNQEPAVIAG 164
DB 133 VMPTISPPPKIAATKYGANITFSGSTSTEREAVTREVIEKTCARLVPPYDHPDIL 192
QY 165 TALEVLNQVPL-----VDALVVPVGGGMAVAGIAITIKALKPS----- 203
DB 193 TAALELQROVAASLSSTGTTTTNRRRLNAITPCGGGLLSGTALACSDLSPSDPTG 252
QY 204 -VKVYAAEP--SNADD---CYOSKLGELTPNLHPETIADGVKSSIGLNTWPII--RDL 255
DB 253 PILVGAEPFSGADGCRGYSGTRIESVKSU----TIADGLRPLGAYPWSIYERKL 308
QY 256 VDDVFTVTEDEIKYATQLVWGRMKLLIETAGVALA-AVLSSHQFQTV-----SPEK 309
```

```

Db 309 VAGWYSGEIEIKALRLVYERMKVVVPEPSAVGLVALFNEPRSMYVERGGEGWDLG 368
Qy 310 IVLSGGNVDLTSL 322
   :| | | | | :| |
Db 369 VVFSGGNVLAAL 381

RESULT 4
Q910F5
ID Q910F5 PRELIMINARY; PRT; 320 AA.
AC Q910F5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Probable serine/threonine dehydratase, degradative.
GN PA2683.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004696; AAG06071.1; -.
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
KW Complete proteome.
SQ SEQUENCE 320 AA; 33957 MW; C6F97C6259FE8584 CRC64;

Query Match 31.3%; Score 544; DB 16; Length 320;
Best Local Similarity 39.08; Pred. No. 3.2e-35;
Matches 122; Conservative 62; Mismatches 121; Indels 8; Gaps 4;

Qy 8 SFADVEKAHINIQDSIHLPVLTSSILNQIAGRNLFKCFQKTSFKIRGALNAIRGL 67
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 6 TYDDVIAAARIAAGHNRTPWSSRTLDEELGAEVFEKCNLQRMGAFKFGAFLSRF 65

Qy 68 IPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGASIVY 127
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 66 ---SAEQRAAGVAFSSNGHQAQIALSARLIGIPATIVMPADAPAVKIEATRGYGQVWL 122

Qy 128 CDPSDESREKVTORIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLVDALVVPVGG 187
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 123 YDRYTEQIGRLDAGRHGLTLPYDHPDLVLAGQGTAAKELFEVGPDLDAFPAPLGGG 182

Qy 188 GWVAGIAITIKALPKSVKYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS-IGL 246
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 183 GLLSGCALRALAPACRIKGVPEAGNDGQSLRSGAIV-HIDTPQTLADGAQTQHLGN 241

Qy 247 NTWPIIRDLDVDDVTVTDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVSPEVK 306
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 242 LTFPLIQNVDDIILTADAEVLDMRELAARMKLLVEPTGCLGLAAAAQRKDEL---RGK 298

Qy 307 NVCIVLSGGNVDL 319
   :| | | | | | | |
Db 299 RVGILLSGGNVDL 311

RESULT 5
Q99050
ID Q99050 PRELIMINARY; PRT; 346 AA.
AC Q99050;

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SAI271 protein (Threonine deaminase IlvA homolog).
GN SAV1438 OR SAI271.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57600.1; -.
DR EMBL; AP003133; BAB42531.1; -.
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilvA_1cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Complete proteome.
SQ SEQUENCE 346 AA; 37147 MW; 4435206B7E89CC10 CRC64;

Query Match 30.7%; Score 535; DB 16; Length 346;
Best Local Similarity 37.8%; Pred. No. 1.9e-34;
Matches 122; Conservative 64; Mismatches 131; Indels 6; Gaps 4;

Qy 5 YCISFADVEKAHINIQDSIHLPVLTSSILNQ- IAGRNLFKCFQKTSFKIRGALNA 63
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 12 HIVSLGDIIEAKSIKPIRTPLIKSMYLSQNTKGNVYKLENMQTSGFKFRGASNK 71

Qy 64 IRGLIPDTPPEKPAVTHSSNGHQALTYAAKLEGIPAYIVVQTPAPNCKKLAIQAYGA 123
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 72 INHL---SDEQKAGLIICASAGNHAQGVALTAKLLGIDATIVMPETAPIAKONATKGXGA 128

Qy 124 SIIVYCDPSDESREKVTORIMQETEGILVHPNQEPAVIAGQGTIALEVLNQPLVDALVVP 183
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 129 KVIILKGNFNETRLYMEELAKENGMTIVHPYDDKRFVMAGQGTIGLEILDDIWNVTIVP 188

Qy 184 VGGGGMVAGIAITIKALPKSVKYAAEPSNADDCYQSKLKGELTPNLHPPTIADGVKSS 243
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 189 VGGGGLIAGIATALKSNFSLIITGVQAEVHGMAESFYKRALTEH-REDSTIADGCDVK 247

Qy 244 I-GLNTWPIIRDLDVDDVTVTDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQTVS 302
   :| | | | | :| | | | | :| | | | | :| | | | | :| | | | | :| |
Db 248 VPGEKTEVVVKHLVDEFTLVSEEEIEHAMQDLQRAKIITEGAGALPTAAILSGKIDKKW 307

Qy 303 PEVKNVCIVLSGGNVDLTSLNW 325
   :| | | | | | | |
Db 308 LEGKNVVALVSGGNVDLTRVSGV 330

RESULT 6
Q9XAA4
ID Q9XAA4 PRELIMINARY; PRT; 325 AA.
AC Q9XAA4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative threonine dehydratase.
GN SC00821 OR SCF43A.11C.

```

OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RA "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RA "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL096837; CAB48898.1;  
 DR HSSP: P04968; 1TDJ.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydratase.  
 DR Pfam: PF00291; PALP; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 SQ SEQUENCE 325 AA; 33965 MW; 4E1D62D340489A7 CRC64;  
 Query Match 30.5%; Score 530; DB 16; Length 325;  
 Best Local Similarity 39.3%; Pred. No. 4.2e-3;  
 Matches 123; Conservative 54; Mismatches 128; Indels 8; Gaps 4;  
 QY 7 ISFADVERKAHNIQDSIHLTPVLTSIIINOTAGRNLFKCELFQKTSFKIRGALNAIRG 66  
 Db VLTDDVRSAAARIKVAHRTPLVRSRTLDALVGAELVKCNQQRVGAFFRGAYNAASR 67  
 QY 67 LIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYGASIV 126  
 Db L---TPEQLARGIAAYSSGNHAQVALAARELGITAVIVMPEDAPPSKRDATAGV 124  
 QY 127 YCDPSDESREKVTQIMQETGILVHPNQEPVAVIAGQGTIALEVNQVPLVDALVVPVGG 186  
 Db TYDRYTGDRVAEALADRLGTLIPPEYHPVHVIAGQGTAALEVEETGELDIALV 184  
 QY 187 GGMVAGIAITTKALKPSVKVYAAPSNAADCCYQSKLKGELTPNLHPPTIADG-VKSSIG 245  
 Db GGLIAGSATYAKALHPGHRVIGVEPADGDTTRSLAAGRRV-SVPVPTIADGALPPG 243  
 QY 246 LNTWPIIRDLVDVFTVTEDEIKYATQLVWGMKLLIEPTAGVALAAVLSQHFQTVSPEV 305  
 Db ELTFSLNRRLLDGLVLSDDSDIADAMRFAERLTKVLEPSPGATPLAALLNGRIDALP--- 300  
 QY 306 KNCVIVLSGGNVD 318  
 Db RVGVILSGGNVD 313

RESULT 7  
 Q9WYJ1  
 ID Q9WYJ1 PRELIMINARY; PRT: 401 AA.  
 AC Q9WYJ1;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Threonine dehydratase catabolic.  
 GN TM0356.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickley E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett K.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima.";  
 RL Nature 393:323-329(1999).  
 DR EMBL: AE001716; AAD35443.1;  
 DR HSSP: P04968; 1TDJ.  
 DR TIGR: TM0356;  
 DR InterPro: IPR002912; ACT.  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydratase.  
 DR Pfam: PF01842; ACT; 1.  
 DR Pfam: PF00291; PALP; 1.  
 DR TIGRFAMS: TIGR011127; ilva\_lcterm; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 401 AA; 43110 MW; 4D757A8F93786180 CRC64;  
 Query Match 30.0%; Score 522.5; DB 16; Length 401;  
 Best Local Similarity 35.8%; Pred. No. 2.3e-3;  
 Matches 114; Conservative 72; Mismatches 123; Indels 9; Gaps 4;  
 QY 7 ISFADVERKAHNIQDSIHLTPVLTSIIINOTAGRNLFKCELFQKTSFKIRGALNAIRG 66  
 Db ITLEDIKEAQRTLNKNVHRTALAYSSVLSVEVTGGIYLUKMNLOKTSFKIRGAYNKIAH 61  
 QY 67 LIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIQAYGASIV 126  
 Db L---SEERKRGVVAASAGNHAQVALAARIGIPATIVMPRYAPLSKITKTRNLGAQVI 118  
 QY 127 YCDPSDESREKVTQIMQETGILVHPNQEPVAVIAGQGTIALEVNQVPLVDALVVPVGG 186  
 Db LEGNIFDEAYEAALRIQEKTCGAVFVHPFNDPHVIAGQGTIGLEIMEDLPDVEVVVPPVGG 178  
 QY 187 GGMVAGIAITTKALKPSVKVYAAPSNAADCCYQSKLKGELTPNLHPPTIADG-VKSSIG 245  
 Db GGLISGVSATKSNPEVKVIGVOTENMPSMIASLRGR-AERVEGKPTLADGIAVKPKP 237  
 QY 246 LNTWPIIRDLVDVFTVTEDEIKYATQLVWGMKLLIEPTAGVALAAVLSQHFQTVSPEV 305  
 Db DTETELVKKYDEVAVVAVNEEIEADALLFLLEQAKVVAAGAGVAAVLNK----LDVKG 293  
 QY 306 KNCVIVLSGGNVD 323  
 Db KKAIVISGGNIDVNMID 311  
 RESULT 8  
 Q8UBW5  
 ID Q8UBW5 PRELIMINARY; PRT: 324 AA.  
 AC Q8UBW5;







	Matches	116: Conservative	67: Mismatches	112: Indels	20: Gaps	57:
Qy	14	KAHINTQDSIHLTPVLTSSILNQIAGLNRLFFKCELFQKTGSKIRGALNAIRGLIPDTPE	73			
Db	11	KAKKNIEDVVIKPLIYSEVFSRKSGNQVMKCNELQLTGAYKIRGALNKIRSL	--SDE	67		
Qy	74	EKPAVNVTHSSGHHGQALTAAKLEGIPAVIVPQTPAPNCKKLAIQAYGASI	-----VYC	128		
Db	68	EKSGKGVVCSAGNHAOGVAFAASQANVKSTIVMPKTTPLIKIQSTKDLGGNVLSGYVYD	127			
Qy	129	DPDSESRKVTYQIMQETEGILVHPNOEPVAVIAGQGTIALEVLNQVPLVDALVVPVCGGG	188			
Db	128	DATNEAK-----RIEQGGALFTHPFNDIDVCGQGTVALEIFEDLNDVDIILCPIGGG	182			
Qy	189	MVAGIAITIKALPKSVKVAAPESNADDCYQSKLKGELTPNLHPETIADGVK--SSIGLN	247			
Db	183	LISGVTLAAKALNPVKVIGVQAEANAMVSKFAGEIIA-LDAVDIADGIANKRPGDL	241			
Qy	248	TWPIIRDLVDVFTVTEDEIKYATQLVMGRMKLLIEPTAGVALAAVLSSHQFTVSPEYKN	307			
Db	242	TFKFIKEYVDIIITVSDHEIVEAFFTLSEKHKLAAEASGAVSLAASAK-----LNCKDN	296			
Qy	308	VCIVLSGGNVDLTSL	322			
Db	297	IVSGVSGNIDMVTI	311			

```

RESULT 13
Q980P1 PRELIMINARY; PRT; 405 AA.
AC Q980P1;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Threonine dehydratase catabolic (threonine deaminase) (tdcB) (EC
DE 4.2.1.16).
DE TDCB OR SS00248.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21322296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awaysz M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikmgoc Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duquet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus p2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006661; AAK40587.1;
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF01842; ACT; 1.
DR TIGRFAMS; TIGR01127; ilva_1cterm; 1.
DR TIGRFAMS; TIGR01127; ilva_1cterm; 1.
SW Lysase; Complete proteome.
SQ SEQUENCE 405 AA; 44363 MW; F1266A940110B062 CRC64;

```

	Query Match	28.7%	Score 499;	DB 17;	Length 405;
	Best Local Similarity	39.1%	Pred. No. 1.7e-31;		
	Matches 126;	Conservative 60;	Mismatches 124;	Indels 12;	Gaps 8;
Qy	5	YCISFADVERKAHINQDSIHLTPVLVTSSILNQIAGRNLFKFCELFQKTSFKIRGALNAI	64		
Db	3	YLYFDRIRAKKEIKYVHTPIDYSTTTSRIINAKVILKLEMLQKTSFKVRGAFNKL	62		
Qy	65	RGLPDTPPEKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPQATPNCKKLAIQAYGAS	124		

Db	63	LSL---KEEEKKNCVIAVSAGNHQAQVAYAASTINIKSTIVMPETAPASKYATKSYGAE	119
Qy	125	IV-YCDPSDESREKVTORIMQETEGILVHPNQEPAVIAQOCTIALEVLNQVPLVDALVVP	183
Db	120	VWLYGKYLHESMKK-AEELIQTGLIFVHPYSLDLDVITGOCTIGLELYDIEP--DYVITP	176
Qy	184	VGGGWAGTAITTKALKPSVKVYAAEPSNADDCYQSKLKGELPNLHPPTIADG--VK	241
Db	177	IGGGGLSGISIALKYRFPNVKTIIGVOSSSSPSMKVSKDLGRLV-EIEPSYSIADGILVK	235
Qy	242	SSIGLNPWPIRLDVDVFTVTEDEIKYATOLVGRMKLLIETPAGVALAAVLSQHFOTV	301
Db	236	SPEL-TFSINELVDLIVDDDEIEAEIAIVLLERSKTLAEGAGAAALASLISGKVK-V	293
Qy	302	SPEYKNVCIVLSGGNVDLTSLN	323
Db	294	NGIDKKVISLVSGGNIDLULLS	315
RESULT	14		
Q9A307			
ID	O9A307	PRELIMINARY:	PRT: 325 AA.

```

RESULT 14
Q9A307 PRELIMINARY; PRT; 325 AA.
Q9A307;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Threonine dehydratase.
GN CC3105.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
SEQUENCE FROM N.A.
RP RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR HSP: P04968; 1TDJ.
DR HSP: CC3105.
DR TGR: CC3105.
DR InterPro: IPR001926; B6_enzyme_beta.
DR Pfam: PF00291; PALP; 1.
DR Complete proteome.
SQ SEQUENCE 325 AA; 33709 MW; E71D0AC7033115E4 CRC64;

```

Query Match	28.6%	Score 498.5	DB 16	Length 325
Best Local Similarity	36.3%	Pred. No. 1.4e-31		
Matches 115	Conservative 58	Mismatches 133	Indels 11	Gaps 4
Qy	7	ISFADVEKAHINIODSIHLTPVLTSSILNQIAGRNLPFKCELFQKTSFKIRGALNAIRG	66	
Db	3	VSLADIQAAVRLKGSAREVPLEIESPALNDRLGGRIFLKPETLQRAGAFKFGAYNRLSQ	62	
Qy	67	LIPDTPPEKPKAVYTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIQAYGASTV	126	
Db	63	L---SDEEKARGVVAFSGNHQAQVLAARLLGVPAIIVMPSDSPSVKVEGTRGFCADIR	119	
Qy	127	YCDPSDESREKVTORIMOETEGILVHPNQEPVIAVAGQCTIALEVLNQV----	182	
Db	120	FYDRFTEDRVAIADQIAERCCVVVPSYDDPHIITAGQGTVGLEIVQAQAAGATLDRLIC	179	
Qy	183	PVGGGVMVAGIATIKALKPSKVYAAEPSNADDCYQSKLKGELTPNLHPPTETADGVKS	242	
Db	180	CVGGGLIAGTSTAVKALSPATEITWVEPAGFDETRRSLESGRRRTIDKARSICDALLT	239	

```
QY 243 SI-GLNTWPIIROLVDDVFTVTDEIKYATOLVWGRMKLLIEPTAGVALAAVLSQLHFTV 301
Db 240 PIPGDLTWPINOKNLGSLVAAVTADEAEAMRYAFSTILKLVPEGGCVALTAAITGKVDVA 299
QY 302 SPEVKNVCIVLSGNDV 318
Db 300 G---KTVAIVLSGNDV 313

RESULT 15
Q9HNH6 PRELIMINARY; PRT; 495 AA.
AC Q9HNH6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Threonine dehydratase.
GN ILDA OR VNG2100G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005099; AGC20244.1; -;
DR HSSP; P04968; ITDJ.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme.beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilva_lcterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 495 AA; 53095 MW; 8F203680AAF93F80 CRC64;

Query Match 28.3%; Score 492.5; DB 17; Length 495;
Best Local Similarity 36.5%; Pred. No. 7.4e-31;
Matches 119; Conservative 59; Mismatches 129; Indels 19; Gaps 7;

QY 7 ISFADVEKAHINIODSIHLTPVLTSSILNOIAGRNLFKCELFQKTSFKIRGALNAIRG 66
Db 87 LSLSDIRDAERVSETHRTPLEYSHTFSDLTGADVRLKLECFQKTSFKIRGATNRIT 146
QY 67 LIPDTPPEKPKAVYTHSSNGHQALTYAAKLEGIPAYIVVPQTPAPNCKKLAIOAGSIV 126
Db 147 LSAD---QQDAGVVVTASAGNHAQVALAASRSQVDSKVMPESAPISKATKSYCAEV 203
QY 127 -----YCDPSDESREKVTQIMQETEG-ILVHPNQEPAVIAGQGTIALEVLNOVPLVDAL 180
Db 204 LHGADYDDAQAHAE-----LEAAEGRTYVHAFDDDEYIMAGQGTGLGLEIAADCPVTDV 257
QY 181 VVPVGGGVAGIATIKALKPSVKVYAAEPSNADDCYOSKLGKELTLPNHPPTIADGV 240
Db 258 VVPIGGGLISGVATALKGELDDVRVIGVOAGCASTVARSLDKQQAQVNH-VDTIADGI 316
QY 241 K-SSIGLNTWPIIROLVDDVFTVTEDEIKYATOLVWGRMKLLIEPTAGVALAAVLSQLHFTQ 299
Db 317 AVROVGAQTFPVIQEHVDEWTVVSDDEIATALLVLLERCKTLVEGAGATAAVALLEDKFE 376
QY 300 TVSPVEKNVCIVLSGNDVLSLNV 325
```

Db 377 YADGE--TIVPALCGGNIDNLTLTV 400

Search completed: June 24, 2003, 06:21:09

Job time : 45.4389 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:35:44 ; Search time 45.9323 Seconds  
(without alignments)  
983.448 Million cell updates/sec

Title: US-09-889-609B-8  
Perfect score: 1740  
Sequence: 1 MCAQYCISPADVEKAHINIQ.....TSLNWMVGOARPAFYQTVSV 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1740	100.0	339	21	AA807731 Amino acid sequenc
2	1582.5	90.9	340	21	AA807734 Amino acid sequenc
3	1582.5	90.9	340	22	AA009124 Human Serine Racem
4	1582.5	90.9	340	22	AA508342 Human pyridoxal-ph
5	1582.5	90.9	340	22	AA594477 Human protein sequ
6	1578.5	90.7	340	22	AA50262 Human serine racem
7	1574.5	90.5	340	22	AA578808 Serine/threonine d
8	988	56.8	228	23	ABB89713 Human polypeptide
9	453	26.0	469	22	ABB63708 Drosophila melanog
10	435.5	25.0	1181	22	ABG24298 Novel human diagno

11	371	21.3	502	20	AA532941 Mutant threonine d
12	371	21.3	502	20	AA505705 Feedback insensiti
13	371	21.3	532	20	AA532943 Mutant threonine d
14	371	21.3	532	20	AA505707 Feedback insensiti
15	371	21.3	539	20	AA532942 Mutant threonine d
16	371	21.3	539	20	AA505706 Feedback insensiti
17	371	21.3	545	20	AA532947 Mutant threonine d
18	371	21.3	545	20	AA505711 Feedback insensiti
19	371	21.3	592	20	AA532939 Mutant threonine d
20	371	21.3	592	20	AA532951 Wild type threonin
21	371	21.3	592	20	AA505702 Arabidopsis wildt
22	371	21.3	592	20	AA505703 Feedback insensiti
23	371	21.3	600	20	AA532952 Mutant threonine d
24	371	21.3	609	20	AA532940 Mutant threonine d
25	371	21.3	609	20	AA505704 Feedback insensiti
26	366	21.0	590	20	AA532950 Mutant threonine d
27	366	21.0	592	20	AA532948 Mutant threonine d
28	365	21.0	424	22	AA581976 S. epidermidis ope
29	365	21.0	424	23	ABP39031 Staphylococcus epi
30	362.5	20.8	416	23	ABB54559 Lactococcus lactis
31	359.5	20.7	441	15	AA54223 L.lactis branched
32	357.5	20.5	310	22	AA590837 C glutamicum prote
33	357.5	20.5	310	22	AA579703 Corynebacterium gl
34	344.5	19.8	422	23	ABB48174 Listeria monocytog
35	325	18.7	621	22	ABG30299 Novel human diagno
36	311	17.9	316	22	ABB66014 Drosophila melanog
37	307	17.6	436	22	AA592074 C glutamicum prote
38	305	17.5	436	16	AA594590 Threonine dehydrat
39	304	17.5	423	22	AA579762 Corynebacterium gl
40	303	17.4	436	16	AA594586 Threonine dehydrat
41	302	17.4	436	16	AA594687 Threonine dehydrat
42	302	17.4	436	16	AA594688 Threonine dehydrat
43	300	17.2	436	16	AA594689 Threonine dehydrat
44	300	17.2	436	16	AA594691 Threonine dehydrat
45	282.5	16.2	340	22	AAU23764 Novel human enzyme

ALIGNMENTS

RESULT 1  
AA807731  
ID AAB07731 standard; Protein; 339 AA.  
XX AC  
XX AAB07731:  
DT 07-NOV-2000 (first entry)  
XX Amino acid sequence of a murine serine racemase polypeptide.  
DE Serine racemase; N-methyl-D-aspartate receptor; neural death;  
XX neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease.  
KW  
XX Mus musculus.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 337  
FT /note= "Ser encoded by T in AAA59294"  
XX  
XX WO200043526-A1.  
XX  
XX 27-JUL-2000.  
XX  
XX 18-JAN-2000; 2000WO-US00938.  
XX  
XX 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX (UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
PA Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
XX  
PI



XX Human; serine racemase; Parkinson's disease; Huntington's disease;  
KW anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia;  
KW schizophrenia; transgenic animal; chromosome 17p13; immunogen.  
XX Homo sapiens.  
OS WO200175144-A1.  
PN 11-OCT-2001.  
XX 02-APR-2001; 2001WO-US10662.  
PF 04-APR-2000; 2000US-194451P.  
PR (MERI ) MERCK & CO INC.  
XX Connolly T, Liu Y, Xia M;  
PI WPI; 2001-656991/75.  
XX N-PSDB; AAS15217.  
DR New recombinant serine racemase polypeptide, useful in assays for  
DR identifying compounds that alter enzyme activity (e.g. including  
DR compounds that inhibit or stimulate enzyme activity) or in generating  
DR antibodies against the protein  
XX Claim 7; Page 14; 43pp; English.  
XX The invention relates to a recombinant polynucleotide encoding a  
CC human serine racemase, vectors containing it, host cells expressing the  
CC racemase, methods of identifying inhibitors of serine racemase and a  
CC transgenic animal lacking a functional endogenous serine racemase  
CC comprising the human serine racemase of the invention. The racemase  
CC protein is useful in assays to identify compounds that inhibit or  
CC stimulate enzyme activity, in the generation of antibodies against the  
CC protein, and in structural studies of the protein and  
CC structure/function relationships of the protein. Biologically active  
CC polypeptide have diagnostic, therapeutic or prophylactic uses (e.g.  
CC for neurological diseases such as Parkinson's and Huntington's disease,  
CC anxiety, glaucoma, stroke, hyperalgesia, pain, spinocerebellar ataxia and  
CC schizophrenia), and would be useful for screening for modulators and/or  
CC inhibitors of serine racemase function. The polynucleotides are useful  
CC as probes for the specific detection of the presence of a polynucleotide  
CC encoding a serine racemase protein, and as primers for nucleic acid  
CC amplification based assays for the detection of polynucleotides encoding  
CC serine racemase protein. The transgenic animal is useful for the study of  
CC the tissue and temporal specific expression or activity of the serine  
CC racemase gene in an animal. The gene for human serine racemase is  
CC located on chromosome 17p13. The present sequence represents human serine  
CC racemase.  
XX Sequence 340 AA;  
SQ

Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINTQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
DB 1 MCAQYICISFADVEKAHINTQDSIHLTPVLTSSILNQIAGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120  
DB 61 LNAIRSLVPDAERKPAVVTHSSGNHGQALTYAAKLEGIPAVIVVQTPAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNQEPVIAAGOTTALEVLNOVPLVDAL 180  
DB 121 YGASIVYCDPSDESRENKRVTEETEGIMVHPNQEPVIAAGOTTALEVLNOVPLVDAL 180  
QY 181 VVPVGGGVMAGIATVTKALPKSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGV 240  
DB 181 VVPVGGGVMAGIATVTKALPKSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGV 240

DB 181 VVPVGGGVMAGIATVTKALPKSVKVYAAEPSNADDCYOSKLGKLMPLNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKATQTLVWGRMKLLIETPTAGVALAVALSOHFQT 300  
DB 241 KSSIGLNTWPIIRDLVDDVFTVTEDEIKATQTLVWGRMKLLIETPTAGVALAVALSOHFQT 300  
QY 301 VSPEVKNVCIVLSSGGNVDLT-SLNWVGQAERPAPYQTVSV 339  
DB 301 VSPEVKNVCIVLSSGGNVDLTSSITWVWQAERPASYQSVSV 340  
RESULT 4  
AAE08342  
ID AAE08342 standard; Protein; 340 AA.  
XX AC AAE08342;  
XX 15-NOV-2001 (first entry)  
XX Human pyridoxal-phosphate dependent enzyme 22406 protein.  
XX Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;  
KW anticonvulsant; cerebroprotective; cardiant; vasotropic; gene therapy;  
KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;  
KW behavioural change; neurodegenerative disorder; schizophrenia; atresia;  
KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;  
KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;  
KW tumour; rhabdomyosarcoma; dermal fibroblast disorder.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Modified-site 8..11  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Domain 19..315  
FT /note= "Pyridoxal-phosphate dependent enzyme family  
FT domain"  
FT Modified-site 38..40  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 47..60  
FT /note= "Serine/threonine dehydratases pyridoxal-phosphate  
FT attachment site"  
FT Modified-site 54..56  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 59..64  
FT /label= N\_myristoylation\_site  
FT Modified-site 88..93  
FT /label= N\_myristoylation\_site  
FT Modified-site 109..112  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 140..143  
FT /note= "cAMP and cGMP-dependent protein kinase  
FT phosphorylation site"  
FT Domain 176..197  
FT /label= Transmembrane\_domain  
FT Modified-site 187..192  
FT /label= N\_myristoylation\_site  
FT Modified-site 196..198  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 203..205  
FT /label= Protein\_kinase\_C\_phosphorylation\_site  
FT Modified-site 212..215  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 235..238  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 239..244  
FT /label= N\_myristoylation\_site  
FT Modified-site 261..264  
FT /label= Casein\_kinase\_II\_phosphorylation\_site  
FT Modified-site 287..292  
FT /label= N\_myristoylation\_site  
FT Domain 308..326  
FT /label= Transmembrane\_domain

XX WO200160987-A1.  
PN XX  
PD XX  
XX XX  
XX 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US05365.  
XX 17-FEB-2000; 2000US-0183208.  
XX (MILL-) MILLENNIUM PHARM INC.  
PA Meyers RA, Rudolph-Owen LA;  
XX WPI; 2001-529909/58.  
PI N-PSDB; AAD14461.  
DR Novel polypeptide of the human pyridoxal phosphate dependent family  
PT useful in screening and detection assays and for treatment, e.g. of  
PT epilepsy and Alzheimer's  
XX Claim 8; Fig 1; 121pp; English.  
PS The present sequence is human pyridoxal phosphate dependent enzyme  
XX 22406 which is a serine racemase. Human 22406 gene is located on  
CC chromosome 17 between D17S849 and D17S796. The protein 22406  
CC is a modulator of D-serine. D serine has been shown to modify  
CC behavioural changes associated with learning, memory and convulsions.  
CC Human 22406 and compounds that modulate the expression or activity are  
CC used to treat or diagnose neurodegenerative disorders including  
CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
CC epilepsy and prevent damage from stroke as well as cardiac (heart  
CC failure, rheumatic heart failure) and circulatory disorders, liver  
CC disorders (hepatic injury, jaundice), lung disorders, prostate  
CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
CC therapy.  
XX XX  
SQ Sequence 340 AA;  
Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINIQDSTHLPVLTSSILNOIAGRNLFKCELFQKTSFKIRGA 60  
Db 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
Db 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTORIMQETGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCDPSDESRENAKRVTEETEGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVAGTAITIKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
Db 181 VVPVGGGVAGTAITVFKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDVFTTVEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFOT 300  
Db 241 KSSIGLNTWPIIRDLVDDIFTTVEDEIKATQLVWGRMKLLIETPTAGVAAVLSHQFOT 300  
QY 301 VSPVKNVICVLSSGGNVDLT-SLNWVGQARPAVYQTVSV 339  
Db 301 VSPVKNVICVLSSGGNVDLTSSITWVQARERASVQSVSV 340  
RESULT 5  
ID AAB94477  
XX AAB94477 standard; Protein; 340 AA.

AC AAB94477;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:15149.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 15149; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX Sequence 340 AA;  
SQ  
Query Match 90.9%; Score 1582.5; DB 22; Length 340;  
Best Local Similarity 89.7%; Pred. No. 3.7e-153;  
Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;  
QY 1 MCAQYICISFADVEKAHINIQDSTHLPVLTSSILNOIAGRNLFKCELFQKTSFKIRGA 60  
Db 1 MCAQYICISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGA 60  
QY 61 LNAIRGLIPDTPPEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
Db 61 LNAVRLVDPALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPNCKKLAIOA 120  
QY 121 YGASIVYCDPSDESREKVTORIMQETGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
Db 121 YGASIVYCDPSDESRENAKRVTEETEGILVHPNQEPVAVTAGOGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGVAGTAITIKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
Db 181 VVPVGGGVAGTAITVFKALKPSKVYAAEPSNADDCYOSKLGELTPNLHPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLVDVFTTVEDEIKYATQLVWGRMKLLIETPTAGVALAVALSHQFOT 300  
Db 241 KSSIGLNTWPIIRDLVDDIFTTVEDEIKATQLVWGRMKLLIETPTAGVAAVLSHQFOT 300  
QY 301 VSPVKNVICVLSSGGNVDLT-SLNWVGQARPAVYQTVSV 339  
Db 301 VSPVKNVICVLSSGGNVDLTSSITWVQARERASVQSVSV 340  
RESULT 5  
ID AAB94477  
XX AAB94477 standard; Protein; 340 AA.





```

XX WPI; 2001-530468/59.
DR N-PSDB; AAI65020.
XX
PT Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding
PT this polypeptide -
XX
PS Claim 1; Page 27 (Disclosure); 34pp; Chinese.
XX
CC The present sequence is the protein sequence for serine/threonine
CC dehydrase 37. The dehydrase and its coding sequence are useful for
CC treating diseases e.g. cancer and HIV infection.
XX
SQ Sequence 340 AA;
Query Match 90.5%; Score 1574.5; DB 22; Length 340;
Best Local Similarity 89.4%; Pred. No. 2.4e-152;
Matches 304; Conservative 19; Mismatches 16; Indels 1; Gaps 1;
QY 1 MCAQYCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFKGTGSKIRGA 60
DB 1 MCAQYCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFKGTGSKIRGA 60
QY 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120
QY 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNQEPAVIAGOGTIALEVLNOVPLVDAL 180
DB 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNQEPAVIAGOGTIALEVLNOVPLVDAL 180
QY 181 VVPVGGGMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTADGV 240
DB 181 VVPVGGGMVAGIATIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPTADGV 240
QY 241 KSSIGLNTWPIIRDLDVDDVTVEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQT 300
DB 241 KSSIGLNTWPIIRDLDVDDVTVEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQT 300
QY 301 VSPVKVNCIVLGGNNVDLT-SLNWVGOAERPAPYQTVSV 339
DB 301 VSSEVRNICIVLGGNNVDLTSSITWVKQERPAVSQSVSV 340
RESULT 8
ABB89713
ID ABB89713 standard; Protein; 228 AA.
XX
AC ABB89713;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 2089.
XX
KW Cytostatic; immunosuppressive; n0otropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;

```

---

```

XX WPI; 2002-122018/16.
DR N-PSDB; ABL90122.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 11; SEQ ID NO 2089; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 228 AA;
Query Match 56.8%; Score 988; DB 23; Length 228;
Best Local Similarity 89.6%; Pred. No. 1.5e-92;
Matches 189; Conservative 12; Mismatches 10; Indels 0; Gaps 0;
QY 1 MCAQYCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFKGTGSKIRGA 60
DB 1 MCAQYCISFADVEKAHINIQDSIHLTPVLTSSILNQIAGRNLFKCELFKGTGSKIRGA 60
QY 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120
DB 61 LNAIRGLIPTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIOA 120
QY 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNQEPAVIAGOGTIALEVLNOVPLVDAL 180
DB 121 YGASIVYCDPSDESREKVTQRIHQETEGILVHPNQEPAVIAGOGTIALEVLNOVPLVDAL 180
QY 181 VVPVGGGMVAGIATIKALKPSVKVYAAEP 211
DB 181 VVPVGGGMVAGIATIKALKPSVKVYAXXP 211
RESULT 9
ABB63708
ID ABB63708 standard; Protein; 469 AA.
XX
AC ABB63708;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 17916.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.

```

OS	Homo sapiens.
XX	
PX	WQ200175067-A2.
NN	
PD	11-OCT-2001.
PP	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PA	(HYSE-) HYSEQ INC.
PI	
DR	Dzmanac RT, Liu C, Tang YT;
PS	NPI; 2001-639362/73.
XX	WI-PSDB; AAS88485.
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
CC	Claim 20; SEQ ID NO 54657; 103pp: English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (III) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1181 AA;
	Query Match 25.0%; Score 435.5; DB 22; Length 1181;
	Best Local Similarity 37.8%; Pred. No. 6.7e-35;
	Matches 105; Conservative 55; Mismatches 105; Indels 13; Gaps 6;
QY	42 LFFKCELFQKTGSFKIRGALNATGLPDPPEPKAVVTHSSNGHGAAITYAAKLEGIP 101
Db	: :   :   :           :   :   :   :   :   :   :   :   :   :
	15 IFUKFENMQRTGSFKIRGAFAKNLSSL---TDAERKKGVACAGNHAGVSLSCAMLGID 71
QY	: :   :   :           :   :   :   :   :   :   :   :   :   :
Db	102 AYIVVPQTAPCNCKLATIQAYGCASIT-YCDPSDESREKVQRIMQETEGILVHPNOEPAVI 160
	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	72 GKVVMPKGAPKSVAATPCDYSAEVVLHGDNFNDTIKAVEISEVMEAR-IFIPPVDPKVI 130
QY	161 AGCGTTALEVLNQPLDALVVPVGGMVAGIAITTKALKPSVKVYAAREFSNADDCVQS 220
	: :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	131 AGGGTGLEITMEDLYDDNVNIVFGGGGLAGIAVAIKSINPTIRIGVQSVGHMAAS 190
QY	221 KLKGELTPNLHPPETIADGVK-SSIGLNTWPRI RDLVDDVFTVFTEDEKYATQLVWGMRK 279
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	191 FHSGEITH-RTTGTADGCDVSRRPGNLTVIEIVRELVDVVLDVSEDELRSNMIALIQRNK 249
QY	280 LLIEPTAGVALAAVLDSQH-----FQTSVEPVKNVCIV 311
	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	250 VVTEGAGALCAALLSEAARGYEIFEKEBECRKLIV 287

QY	180	LVPVPGGGMVAGIAITIKALPKSPVKVYAAEPSNADDCYQSKLKGELTPTNLHPPETI	236
DB	173	IFVPVGGGLIAGIAAAYKRVSPVKKIIGVEPADAN-----AMALSLLHHGERVILD	223
QY	237	-----ADGVK-SSIGLNTWPIIRDLDVDDVFTTDEIKYATOLVWGRMKLLLTPTAGVAL	250
DB	224	QVGFADGVAVKVEGETFRISRLNMDGVVLVTRDAICASIKDMFEKRNILLEPAGALAL	283
QY	291	AAVLSSQHFQTPVSPEVKNCIVLSGGGNVDLTSLNWV	325
DB	284	AGA-EAYCKYGYGLKDVNVVAITSGANMFDKLRIV	317
QY	RESULT 12		
AY	AY05705	AY05705	
XX	AY05705	standard; Protein; 502 AA.	
AC	AY05705;		
XX	19-JUL-1999	(first entry)	
DT	Feedback insensitive mature threonine dehydratase/deaminase.		
DE	Threonine dehydratase/deaminase; omr1 gene; feedback inhibition;		
KW	transgenic plant; selectable marker; isoleucine; mutant.		
KW	Arabidopsis thaliana.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
FH	Misc-difference 409	/note= "Arg in wild-type enzyme"	
FT	Misc-difference 454	/note= "Arg in wild-type enzyme"	
FT	Region	396...414	
FT	Region	/note= "regulatory region R4"	
FT	Region	446...464	
FT	Region	/note= "regulatory region R6"	
XX	WO9902656-A1.		
XX	21-JAN-1999.		
XX	10-JUL-1998;	98WO-US14362.	
PF	17-FEB-1998;	98US-0074875.	
XX	10-JUL-1997;	97US-0052096.	
PR	(PURD ) PURDUE RES FOUND.		
XX	Mourad GS;		
PI	WPI; 1999-120860/10.		
XX	DR NP-SDB; AAX25334.		
DR	New sequences encode mutant threonine dehydratase/deaminase - which		
XX	is insensitive to feedback inhibition, useful as a selective marker		
PT	to produce transformed cells resistant to toxic isoleucine analogues		
PT	Disclosure; Page 60-62; 120pp; English.		
PS	This sequence represents an Arabidopsis thaliana mutant mature		
XX	threonine dehydratase/deaminase (TD) which, unlike the wild-type		
CC	enzyme (see AAY05702), is insensitive to feedback inhibition by		
CC	isoleucine. It lacks the chloroplast transit peptide of the		
CC	mutant TD precursor, and is encoded by a polynucleotide (see		
CC	AAX25334) that is one of 9 claimed polynucleotides (see AAX25332-40),		
CC	originally isolated and cloned from A. thaliana mutant line GM11b		
CC	(omr1/omr1), which encode a feedback insensitive TD. These		
CC	polynucleotides can be used to transform a wide variety of plants,		
CC	fungi, bacteria and yeast. The polynucleotides differ from the		
CC	wild-type only by 2 point mutations, which result in R499G and		

Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
WPI: 1999-527375/44.  
N-PSDB; AA211201.

New nucleic acid encoding threonine dehydratase deaminase resistant  
to feedback inhibition, useful as selection marker for cell  
transformation and to impart herbicide resistance

Claim 13; Page 111-114; 194pp; English.

This sequence represents a mutant Arabidopsis thaliana threonine  
dehydratase/deaminase (TD) protein of the invention. The protein is a  
feedback insensitive mutant. The TD DNA sequence is used as molecular  
marker (imparting resistance to toxic structural analogues of isoleucine)  
for selected transformed cells and to produce transformants with  
increased levels of isoleucine (and thus better nutritional value) or of  
intermediates in biosynthesis of isoleucine (e.g. 2-oxobuturate, for  
synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
TD-expressing plants permit use of the isoleucine structural analogues as  
herbicides. The DNA sequences are alternatives for antibiotic resistance  
markers (which are potentially harmful to the environment). Since no  
human analog of TD exists (humans can not synthesize isoleucine), it  
should be safe to use.

Sequence 532 AA;

Query Match 21.3%; Score 371; DB 20; Length 532;  
Best Local Similarity 31.6%; Pred. No. 8e-29;  
Matches 106; Conservative 61; Mismatches 134; Indels 34; Gaps 11;

QY 4 QYCISFADVEKAHINIODSIHLTPVLVTSSTILNCIAGRNLFLKCELFQKTGSKIRGALNA 63  
:  
Db 34 EYLTNLSTKVVDIAIESPLQAKLSKRL-----GVRYLKREDLPQVFSEFKLRGAYNM 88  
:  
QY 64 IRGLIPDTPPEPKAVYTHSSNGHGQALTAAKLEGIPAYIVVPOTAPNCKLIAIQAYGA 123  
:  
Db 89 MVKLPAID---QLAKGVICSSAGNHQAQVALSASKLGCTAIVMPVTTPETIKWAQVENLGA 145  
:  
QY 124 SIV-YCDPDSEREKVTQIMQTEGI-LVHPNQEPAVIAGQGTTALEVLNQV--PLVDA 179  
:  
Db 146 TVVLEFGSDVDAQAHA--KIRAEELGTFTPPDPHDVIAGQGTVGMEITRQAKGPL-HA 202  
:  
QY 180 LVVPGGGGMVAGIATIKALPSVKVYAEPNSADDQCYGSKLKGELTPNLHPPETI--- 236  
:  
Db 203 IFPVGGGGGLIAGIAYVRSPVKLIIGVEPADAN-----AMASLHHGERVILD 253  
:  
QY 237 -----ADGVK-SSIGLTWPRIIDLVDVVFTVFEDEIKYATQLVWGMRMKLLIEPTAGVAL 290  
:  
Db 254 QVGGFADGVAVKEVGEETFRISRNLMDGVVLVTRDALCASIKDMFEKRNIILEPAGALAL 313  
:  
QY 291 AAVLSOHFOTVSPVKNCVICLVSGGNVDLTSLNMV 325  
:  
Db 314 AGA-EAYCKYYGLKDENVVAITSGANNFDKLRIV 347  
: :

RESULT 14  
AAAY05707  
ID AAAY05707 standard; Protein: 532 AA.  
AC AAAY05707;  
DT DT  
XX 19-JUL-1999 (first entry)

Feedback insensitive mature threonine dehydratase/deaminase.  
Threonine dehydratase/deaminase; omr1 gene; feedback inhibition;  
transgenic plant; selectable marker; isoleucine; mutant.

Arabidopsis thaliana.  
OS Synthetic.

FH Key Location/Qualifiers  
 FT Misc-difference 439 /note= "Arg in wild-type enzyme"  
 FT Misc-difference 484 /note= "Arg in wild-type enzyme"  
 FT Region 427..444  
 FT /note= "regulatory region R4"  
 FT Region 476..494  
 FT /note= "regulatory region R6"  
 XX WO9902656-A1.  
 PN 21-JAN-1999.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 PR 10-JUL-1997; 97US-0052096.  
 XX (PURD ) PURDUE RES FOUND.  
 PA Mourad GS;  
 XX WPI; 1999-120860/10.  
 DR N-PSDB; AAX25336.  
 XX New sequences encode mutant threonine dehydratase/deaminase - which  
 PT is insensitive to feedback inhibition, useful as a selective marker  
 PT to produce transformed cells resistant to toxic isoleucine analogues  
 XX Disclosure: Page 65-68; 120pp; English.  
 PS This sequence represents an Arabidopsis thaliana mutant mature  
 CC threonine dehydratase/deaminase (TD) which, unlike the wild-type  
 CC enzyme (see AAY05702), is insensitive to feedback inhibition by  
 CC isoleucine. It lacks the chloroplast transit peptide of the  
 CC mutant TD precursor, and is encoded by a polynucleotide (see  
 CC AAX25336) that is one of 9 claimed polynucleotides (see AAX25332-40),  
 CC originally isolated and cloned from A. thaliana mutant line GM1lb  
 CC (omrl/omrl), which encode a feedback insensitive TD. These  
 CC polynucleotides can be used to transform a wide variety of plants,  
 CC fungi, bacteria and yeast. These polynucleotides differ from the  
 CC wild-type only by 2 point mutations, which result in R499C and  
 CC R554H amino acid substitutions (numbering according to wild-type  
 CC TD) in the R4 and R6 regulatory regions. These forms of TD are not  
 CC only insensitive to feedback inhibition by isoleucine, but are also  
 CC insensitive to structural analogues of isoleucine that are toxic to  
 CC plants and microorganisms which synthesize only wild-type TD.  
 CC Nucleotide sequences encoding mutated forms of TD can therefore be  
 CC used to create cells that are insensitive to compounds normally  
 CC toxic to cells expressing only wild-type TD enzymes, and thus may  
 CC be used to provide a biochemical selectable marker. Transformants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX Sequence 532 AA;  
 SQ  
 Query Match 21.3%; Score 371; DB 20; Length 532;  
 Best Local Similarity 31.6%; Pred. No. 8e-29;  
 Matches 106; Conservative 61; Mismatches 134; Indels 34; Gaps 11;  
 QY 4 CYCTSFADVERAHNIODSIHLTPVLTSSILNQIAGRNLFKCELFQKSGFKIRGALNA 63  
 DB 34 EYLNILSTKYVDIAIESPLQAKLSKRL-----GVRWYLRDELQVFSFKRGAVNM 88  
 QY 64 IRLGIPDTPPEKPAVYVTHSSGNHQALTYAAKLEGIPAYIVVPQTPNCKKLAIQAYGA 123  
 DB 89 MVKLPAD---QLAKGVICSSAGNAGQVALSASKLGCTAVIVMPVTPTEIKWQAVENLGA 145  
 QY 124 SIY-YCDPSDSREKVTORIMQWETGI-LVHPNQEPAVIAQGGTIALEVLNQV--PLVDA 179  
 DB 146 TVVLFSGSDYDQAHA--KIRAEEGTLTFIPFDHPDVIAGQGTVMETIRQAQGPL-HA 202

QY 180 LVVPVGGGMVAGIAITIKALKPSKVYAAEPSNADDCYQSKLKGELTPNLHPPETI--- 236  
 DB 203 IFVPVGGGLIAGIAAYVKRVSPVKIIGVEPADAN-----AMALSUHGGERVILD 253  
 QY 237 -----ADGVK-SSIGLNTWPIIROLVDVFTVEDEIKYATQLVWGRMKLLIETAGVAL 290  
 DB 254 QVGGFADGVAVKEVGEETFRISRNLMGVLVTRDAICASIKDMFEERKRNILEPAGALAL 313  
 QY 291 AAVLSQHFQTVSPVKVNCIVLSGGNVDLTSLNWV 325  
 DB 314 AGA-EAYCKYGLKDVNVVAITSGANMFDKLRIV 347  
 RESULT 15  
 AAY32942  
 ID AAY32942 standard; Protein: 539 AA.  
 XX  
 AC AAY32942;  
 XX 09-NOV-1999 (first entry)  
 DT Mutant threonine dehydratase/deaminase protein sequence.  
 DE  
 XX Threonine dehydratase/deaminase: TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PF 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOMC ) DOW AGROSCIENCES LLC.  
 XX (PURD ) PURDUE RES FOUND.  
 PI Larinna IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; AAZ11200.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Page 106-109; 194pp; English.  
 PS This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX Sequence 539 AA;  
 SQ Query Match 21.3%; Score 371; DB 20; Length 539;  
 Best Local Similarity 31.6%; Pred. No. 8.1e-29;



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:23:16 ; Search time 285 Seconds  
(without alignments)  
8083.489 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgcattt.....atcagtcgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_101002:*			
1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*		
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*		
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*		
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*		
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*		
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*		
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*		
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*		
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*		
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*		
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*		
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*		
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*		
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*		
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*		
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*		
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*		
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*		
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*		
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*		
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*		
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*		
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*		
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1023	100.0	1023	21	AAA59299
2	1023	100.0	1023	22	AA515217
3	1023	100.0	1023	22	AAAD1461
4	1023	100.0	2477	22	AAH16282
5	1021.4	99.8	1336	22	AAI70575
6	1021.4	99.8	2674	22	AAT65020
7	816.6	79.8	1672	21	AAA59300
8	807.8	79.0	1018	21	AAA59294
9	729.2	71.3	848	22	AAH06600

10	645.6	63.1	731	24	ABL90122	Human polynucleoti
11	535	52.3	608	21	AAA59295	N-terminal sequenc
12	337	32.9	509	21	AAA59296	C-terminal sequenc
13	221	21.6	20892	22	ABA15709	Human nervous syst
14	200.2	19.6	861	21	AAAL0869	Human secreted pro
15	139.8	13.7	31096	18	AAV74370	Staphylococcus aur
16	112.4	11.0	583	22	AAH98041	Murine 7-transmemb
17	104	10.2	1447	23	ABL13783	Drosophila melanog
18	99.4	9.7	1189	23	ABL18395	Drosophila melanog
19	96.6	9.4	1509	20	AAZ11199	Mutant threonine d
20	96.6	9.4	1509	20	AAZ25334	Feedback insensiti
21	96.6	9.4	1599	20	AAZ11201	Mutant threonine d
22	96.6	9.4	1599	20	AAZ25336	Feedback insensiti
23	96.6	9.4	1620	20	AAZ11200	Mutant threonine d
24	96.6	9.4	1620	20	AAZ25335	Feedback insensiti
25	96.6	9.4	1638	20	AAZ11205	Mutant threonine d
26	96.6	9.4	1638	20	AAZ25340	Feedback insensiti
27	96.6	9.4	1779	20	AAZ11197	Mutant threonine d
28	96.6	9.4	1779	20	AAZ11209	Wild type threonin
29	96.6	9.4	1779	20	AAZ25331	Arabidopsis wild-t
30	96.6	9.4	1779	20	AAZ25332	Feedback insensiti
31	96.6	9.4	1830	20	AAZ11198	Mutant threonine d
32	96.6	9.4	1830	20	AAZ25333	Feedback insensiti
33	96.6	9.4	2277	20	AAZ11212	Mutant threonine d
34	94	9.2	542	22	AAI83752	Human polynucleoti
35	93.4	9.1	2235	20	AAZ11208	Mutant threonine d
36	93.4	9.1	2241	20	AAZ11206	Mutant threonine d
37	89.4	8.7	1830121	17	AAZ42063	Haemophilus influe
38	86.6	8.5	3550	23	AA588485	DNA encoding novel
39	84.8	8.3	1478	20	AAZ07180	Corn threonine dea
40	77.4	7.6	105184	24	ABK24122	Bacterial artifici
41	75.4	7.4	1866	23	AA582540	DNA encoding novel
42	75.4	7.4	1866	23	AA594486	DNA encoding novel
43	72.4	7.1	922	18	AAZ84000	DNA encoding a thr
44	71.6	7.0	3402	23	ABL18394	Drosophila melanog
45	71.6	7.0	4179	23	ABL13782	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAA59299  
ID AAA59299 standard; DNA; 1023 BP.  
XX  
AC AAA59299;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE DNA encoding a human serine racemase polypeptide.  
XX  
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1023  
FT /tag= a  
FT /product= "serine racemase"  
XX  
PN WO200043526-A1.  
XX  
PD 27-JUL-2000.  
XX  
PF 18-JAN-2000; 2000WO-US00938.  
XX  
PR 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX  
PA (UYUO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
 PI Ferris CD;  
 XX WPI: 2000-482915/42.  
 DR P-PSDB: AAB07734.  
 XX Mammalian serine racemase preparations, used\* to identify modulators  
 PT which can be used to treat diseases associated with  
 PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
 PT  
 XX Claim 19; Page 49-50; 54pp; English.  
 XX The present sequence encodes a mammalian serine racemase, which has  
 CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
 CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
 CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
 CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
 CC modulators, which can be used in the treatment of acute or chronic  
 CC neuronal death or dysfunction mediated by overactivation of N-methyl-D-  
 CC aspartate (NMDA) receptors. Overactivation of the receptors is  
 CC associated with Parkinson's disease, Huntington's disease, motor neurone  
 CC disease and Alzheimer's disease.  
 XX  
 SO Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;  
 Query Match 100.0%; Score 1023; DB 21; Length 1023;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-310;  
 Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGCTCAGTATTCATCTCTCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 DB 1 ATGTGCTCAGTATTCATCTCTCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 QY 61 GATTCTATCACCTCACACGAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
 DB 61 GATTCTATCACCTCACACGAGTCTAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
 QY 121 AATCTTTTCTCAATGTGAACCTCTCCAGAAACAGGATCTTTTAAGATTCTGGTGCT 180  
 DB 121 AATCTTTTCTCAATGTGAACCTCTCCAGAAACAGGATCTTTTAAGATTCTGGTGCT 180  
 QY 181 CTCGAATGCCGTCAGAAAGCTTGGTTCCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 240  
 DB 181 CTCGAATGCCGTCAGAAAGCTTGGTTCCTGATGCTTTAGAAAGGAGCCGAAAGCTGTGTT 240  
 QY 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300  
 DB 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTCACCTATGCTGCCAAATTTGGAAGGAAT 300  
 QY 301 CCTGCTTATATTGTGTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
 DB 301 CCTGCTTATATTGTGTGCTCCAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360  
 QY 361 TAGCGAGCGTCAATTTGATATCTGTGAACCTAGTATGATGATGATGATGATGATGATGAT 420  
 DB 361 TAGCGAGCGTCAATTTGATATCTGTGAACCTAGTATGATGATGATGATGATGATGATGAT 420  
 QY 421 AGAGTTACAGAAAGAAACAGAGGATCATGTTGATGATGATGATGATGATGATGATGATGAT 480  
 DB 421 AGAGTTACAGAAAGAAACAGAGGATCATGTTGATGATGATGATGATGATGATGATGATGAT 480  
 QY 481 GCTGGACAGGAGCAATTTGCTGGAAGTGTCTGAACAGGTTCTCTTTGTTGGATGCACTG 540  
 DB 481 GCTGGACAGGAGCAATTTGCTGGAAGTGTCTGAACAGGTTCTCTTTGTTGGATGCACTG 540  
 QY 541 GTGGTACCTTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAAGTAAAGCTCTG 600  
 DB 541 GTGGTACCTTGTAGTGGAGGAGGAATGCTTGTGGAATAGCAATTAAGTAAAGCTCTG 600  
 QY 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATTCAGATCACTGCTACCAGTCC 660  
 DB 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATTCAGATCACTGCTACCAGTCC 660

DB 601 AAACCTAGTGTGAAGTATATGCTGTGAACCTCAATTCAGATCACTGCTACCAGTCC 660  
 QY 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGGTCTC 720  
 DB 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTATCTCCAGAAACCATAGCAGATGGTCTC 720  
 QY 721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATTTATCAGGAGACCTTGTGGATGATATCTTC 780  
 DB 721 AAATCCAGCATTTGGCTTGAACACCTGGCTTATTTATCAGGAGACCTTGTGGATGATATCTTC 780  
 QY 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840  
 DB 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAAACTA 840  
 QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTCTCAACATTTTCAAACT 900  
 DB 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGCTGCTGCTCTCAACATTTTCAAACT 900  
 QY 901 GTTTCCCCAGAGTAAAGAACATTTGATTGCTCAGTGGTGGAAATGTAGAGTTAAACC 960  
 DB 901 GTTTCCCCAGAGTAAAGAACATTTGATTGCTCAGTGGTGGAAATGTAGAGTTAAACC 960  
 QY 961 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1020  
 DB 961 TCCTCCATAACTTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1020  
 QY 1021 TAA 1023  
 DB 1021 TAA 1023

RESULT 2  
 AAS15217  
 ID AAS15217 standard; cDNA; 1023 BP.  
 XX AAS15217;  
 AC AAS15217;  
 XX 16-JAN-2002 (first entry)  
 XX Human cDNA encoding Serine Racemase.  
 DE Human; ss; serine racemase; Parkinson's disease; Huntington's disease;  
 KW anxiety; glaucoma; stroke; hyperalgesia; pain; spinocerebellar ataxia;  
 KW schizophrenia; transgenic animal; chromosome 17p13.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT CDS 1..1023  
 FT /\*tag= a  
 FT /product= "Serine racemase"  
 XX WO200175144-A1.  
 XX 11-OCT-2001.  
 XX 02-APR-2001; 2001WO-US10662.  
 XX 04-APR-2000; 2000US-194451P.  
 XX (MERI ) MERCK & CO INC.  
 XX Connolly T, Liu Y, Xia M;  
 DR WPI: 2001-656991/75.  
 DR P-PSDB: AAU09124.  
 XX New recombinant serine racemase polypeptide, useful in assays for  
 PT identifying compounds that alter enzyme activity (e.g. including  
 PT compounds that inhibit or stimulate enzyme activity) or in generating  
 PT antibodies against the protein  
 XX Claim 1; Page 9; 43pp; English.

XX The invention relates to a recombinant polynucleotide encoding a  
CC human serine racemase, vectors containing it, host cells expressing the  
CC racemase, methods of identifying inhibitors of serine racemase and a  
CC transgenic animal lacking a functional endogenous serine racemase  
CC comprising the human serine racemase of the invention. The racemase  
CC protein is useful in assays to identify compounds that inhibit or  
CC stimulate enzyme activity, in the generation of antibodies against the  
CC protein, and in structural studies of the protein and  
CC structure/function relationships of the protein. Biologically active  
CC fragments, and mutant or polymorphic forms of the serine racemase  
CC polypeptide have diagnostic, therapeutic or prophylactic uses (e.g.  
CC for neurological diseases such as Parkinson's and Huntington's disease,  
CC anxiety, glaucoma, stroke, hyperalgesia, pain, spinocerebellar ataxia and  
CC schizophrenia), and would be useful for screening for modulators and/or  
CC inhibitors of serine racemase function. The polynucleotides are useful  
CC as probes for the specific detection of the presence of a polynucleotide  
CC encoding a serine racemase protein, and as primers for nucleic acid  
CC amplification based assays for the detection of polynucleotides encoding  
CC serine racemase protein. The transgenic animal is useful for the study of  
CC the tissue and temporal specific expression or activity of the serine  
CC racemase gene in an animal. The gene for human serine racemase is  
CC located on chromosome 17p13. The present sequence encodes human serine  
CC racemase.  
XX  
SQ Sequence 1023 BP; 290 A; 220 C; 243 G; 270 T; 0 other;

Query Match 100.0%; Score 1023; DB 22; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 2.3e-310;  
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTGTCAGTATGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
Db 1 ATGTGTCAGTATGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTGCA 60  
Qy 61 GATTCTATCCACCTCACACAGTCTCAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
Db 61 GATTCTATCCACCTCACACAGTCTCAACAGCTCCATTTTGAATCAACTAACAGGGCGC 120  
Qy 121 AATCTTTTCTTCAATGTGAATCTTCCAGAAAACAGATCTTTTAAGATTCGTGGTGCT 180  
Db 121 AATCTTTTCTTCAATGTGAATCTTCCAGAAAACAGATCTTTTAAGATTCGTGGTGCT 180  
Qy 181 CTCAATCCGCTCAGAAGCTTGGTTCCTGATGCTTTAGAAAGAACCGGAAAGCTGTGTT 240  
Db 181 CTCAATCCGCTCAGAAGCTTGGTTCCTGATGCTTTAGAAAGAACCGGAAAGCTGTGTT 240  
Qy 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTACCTATGCTGCCAAATTTGGAAGGAATT 300  
Db 241 ACTCACAGCAGTGAACCATGCGCCAGGCTCTACCTATGCTGCCAAATTTGGAAGGAATT 300  
Qy 301 CCTGCTTATATTGTGGTGGCCACACAGCTCCACACTGTAAACACTTGCATACAGGCC 360  
Db 301 CCTGCTTATATTGTGGTGGCCACACAGCTCCACACTGTAAACACTTGCATACAGGCC 360  
Qy 361 TACGAGGCGTCAATTGTATCTACTGTGAACCTAGTATGATGATGATGATGATGATGATG 420  
Db 361 TACGAGGCGTCAATTGTATCTACTGTGAACCTAGTATGATGATGATGATGATGATGATG 420  
Qy 421 AGAGTTACAGAGAAACAGAGGCGATCATGTTGATATCCCAACAGGAGCCGTGCAGTGATA 480  
Db 421 AGAGTTACAGAGAAACAGAGGCGATCATGTTGATATCCCAACAGGAGCCGTGCAGTGATA 480  
Qy 481 GCTGGAAGGGAACAATTGCCTGGAAGTCTGAACAGGTTCTTTGGTGGATGCACTG 540  
Db 481 GCTGGAAGGGAACAATTGCCTGGAAGTCTGAACAGGTTCTTTGGTGGATGCACTG 540  
Qy 541 GTGTACCTGTAGGTGGAGGAGGAATGCTTCTGGAATAGCAATTAAGTAAAGGCTCTG 600  
Db 541 GTGTACCTGTAGGTGGAGGAGGAATGCTTCTGGAATAGCAATTAAGTAAAGGCTCTG 600  
Qy 601 AAACCTAGTGTGAAGGATATGCTGCTGAACCCCTCAATGCAGATGATGCTACCACTGCC 660  
Db 601 AAACCTAGTGTGAAGGATATGCTGCTGAACCCCTCAATGCAGATGATGCTACCACTGCC 660

Db 601 AAACCTAGTGTGAAGGATATGCTGCTGAACCCCTCAATGCAGATGATGCTACCACTGCC 660  
Qy 661 AAGCTGAAGGGGAACATGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTC 720  
Db 661 AAGCTGAAGGGGAACATGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTC 720  
Qy 721 AAATCCAGCAATTTGGCTTTGAACACCTGGCCTATTATCAGGGACCTTGTGGATGATATCTTC 780  
Db 721 AAATCCAGCAATTTGGCTTTGAACACCTGGCCTATTATCAGGGACCTTGTGGATGATATCTTC 780  
Qy 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACACCTGGTGTGGAGAGGATGAAACTA 840  
Db 781 ACTGTCACAGAGGATGAAATTAAGTGTGCAACACCTGGTGTGGAGAGGATGAAACTA 840  
Qy 841 CTCAATGAACCTACAGCTGGTGGAGTGGCTGCTCTCAACATTTTCAAACT 900  
Db 841 CTCAATGAACCTACAGCTGGTGGAGTGGCTGCTCTCAACATTTTCAAACT 900  
Qy 901 GTTTTCCCAGAAAGTAAAGAACATTTGATTTGCTCAGTGTGGAAATGTAGACTTAACC 960  
Db 901 GTTTTCCCAGAAAGTAAAGAACATTTGATTTGCTCAGTGTGGAAATGTAGACTTAACC 960  
Qy 961 TCCTCCATACTTGGGTGAACAGGCTGAAAAGGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Db 961 TCCTCCATACTTGGGTGAACAGGCTGAAAAGGCCAGCTTCTTATCAGTCTGTTTCTGTT 1020  
Qy 1021 TAA 1023  
Db 1021 TAA 1023

RESULT 3  
AAD14461  
ID AAD14461 standard; cDNA; 1770 BP.  
XX AAD14461;  
AC AAD14461;  
DT 15-NOV-2001 (first entry)  
XX Human pyridoxal-phosphate dependent enzyme 22406 cDNA.  
DE Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;  
XX anticonvulsant; cerebroprotective; cardiac; vasotropic; gene therapy;  
KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;  
KW behavioural change; neurodegenerative disorder; schizophrenia; atresia;  
KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;  
KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;  
XX tumour; rhabdomyosarcoma; dermal fibroblast disorder; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
CDS 69..1091  
FT /\*tag= a  
FT /product= "Human pyridoxal phosphate dependent enzyme"  
FT misc\_feature 69..1088  
FT /\*tag= b  
FT /note= "This region is specifically claimed as  
FT SED ID NO: 3 in claim 1"  
XX WO200160987-A1.  
PN 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US05365.  
XX 17-FEB-2000; 2000US-0183208.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Meyers RA, Rudolph-Owen LA;  
XX WPI; 2001-529909/58.  
DR

DR P-PSDB; AAE08342.

XX Novel polypeptide of the human pyridoxal phosphate dependent family  
 PT useful in screening and detection assays and for treatment, e.g. of  
 PT epilepsy and Alzheimer's

XX Claim 1; Fig 1; 12lpp; English.

PS The present sequence is a cDNA encoding human pyridoxal phosphate  
 CC dependent enzyme 22406 which is a serine racemase. Human 22406 gene is  
 CC located on chromosome 17 between D17S849 and D17S796. The protein 22406  
 CC is a modulator of D-serine. D serine has been shown to modify  
 CC behavioural changes associated with learning, memory and convulsions.  
 CC Human 22046 and compounds that modulate the expression or activity are  
 CC used to treat or diagnose neurodegenerative disorders including  
 CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
 CC epilepsy and prevent damage from stroke as well as cardiac (heart  
 CC failure, rheumatic heart failure) and circulatory disorders, liver  
 CC disorders (hepatic injury, jaundice), lung disorders, prostate  
 CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
 CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
 CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
 CC therapy.

SQ Sequence 1770 BP; 499 A; 379 C; 410 G; 482 T; 0 other;

Query Match 100.0%; Score 1023; DB 22; Length 1770;

Best Local Similarity 100.0%; Pred. No. 3.1e-310;  
 Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGTCAGTATGTCATCTCCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
 Db 69 ATGTTGTCAGTATGTCATCTCCTTCTGATGTTGAAAAGCTCATATCAACATTCGA 128  
 QY 61 GATTCTATCCACTCACACAGTCTACACAGTCCATTTTGAATCACTAACAGGGCGC 120  
 Db 129 GATTCTATCCACTCACACAGTCTACACAGTCCATTTTGAATCACTAACAGGGCGC 188  
 QY 121 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGCT 180  
 Db 189 AATCTTTTCTCAATGTGAATCTTCCAGAAAACAGGATCTTTTAAAGATTCGTGCT 248  
 QY 181 CTCAATCGCGTCAGAAAGCTTGTTGCTGATGCTTTAGAAAAGAGCCGAAAGCTTGTCT 240  
 Db 249 CTCAATCGCGTCAGAAAGCTTGTTGCTGATGCTTTAGAAAAGAGCCGAAAGCTTGTCT 308  
 QY 241 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCTATGCTGCCAATTTGAAGGAATT 300  
 Db 309 ACTCACAGCTGGAACCATGGCCAGGCTCTCACCTATGCTGCCAATTTGAAGGAATT 368  
 QY 301 CTGCTTTATTTGTTGTCGCCCCAGACAGCTCCAGACTGTAAAAAACCTTGCAATACAAGCC 360  
 Db 369 CTGCTTTATTTGTTGTCGCCCCAGACAGCTCCAGACTGTAAAAAACCTTGCAATACAAGCC 428  
 QY 361 TACGAGCGCTCAATGTATCTGTGAACCTAGTATGATGATCCAGAGAAAATTTGCAAAA 420  
 Db 429 TACGAGCGCTCAATGTATCTGTGAACCTAGTATGATGATCCAGAGAAAATTTGCAAAA 488  
 QY 421 ACAGTTACAGAGAAACAGAGGATCATGTCATCCACAGGAGCCCTGCAGTGATA 480  
 Db 489 ACAGTTACAGAGAAACAGAGGATCATGTCATCCACAGGAGCCCTGCAGTGATA 548  
 QY 481 GCTGGACAAGGACAATTCCTGCTGAAAGTGTGAACAGGTTTCCTTTGTTGGATGCACTG 540  
 Db 549 GCTGGACAAGGACAATTCCTGCTGAAAGTGTGAACAGGTTTCCTTTGTTGGATGCACTG 608  
 QY 541 GTGGTACCTGTAGGTGGAGGAGGAATGCTTCTGGAATPAGCAATTTACAGTTAAGGCTCTG 600  
 Db 609 GTGGTACCTGTAGGTGGAGGAGGAATGCTTCTGGAATPAGCAATTTACAGTTAAGGCTCTG 668  
 QY 601 AAACCTAGTGTGAGGATATGCTGCTGAACCCCTCAATGTCAGATGACTGTACCAAGTCC 660  
 Db 669 AAACCTAGTGTGAGGATATGCTGCTGAACCCCTCAATGTCAGATGACTGTACCAAGTCC 728

QY 661 AAGCTGAAGGGAAACTGATGCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTGC 720  
 Db 729 AAGCTGAAGGGAAACTGATGCCAATCTTTATCTCTCCAGAAACCATAGCAGATGGTGC 788  
 QY 721 AAATCCAGCATTTGGCTTGAACACCTGGCCTATTATCAGGGACCTTTGTGGATGATATCTTC 780  
 Db 789 AAATCCAGCATTTGGCTTGAACACCTGGCCTATTATCAGGGACCTTTGTGGATGATATCTTC 848  
 QY 781 ACTGTCACAGAGGATGAATTAAGTGTCCACCCAGCTGGTGTGGGAGAGGATGAACATA 840  
 Db 849 ACTGTCACAGAGGATGAATTAAGTGTCCACCCAGCTGGTGTGGGAGAGGATGAACATA 908  
 QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTGCTCACAACATTTTCAAACT 900  
 Db 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTCTGCTCACAACATTTTCAAACT 968  
 QY 901 GTTCTCCAGAGTAAGAACACATTTTGTATTTGCTCAGTGGTGGAAATGTAGACTTAACC 960  
 Db 969 GTTCTCCAGAGTAAGAACACATTTTGTATTTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
 QY 961 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1020  
 Db 1029 TCCTCCATAACTTGGGTGAAGCAGGCTGAAAGGCCAGCTTCTTATCAGTCTGTTCTGTT 1088  
 QY 1021 TAA 1023  
 Db 1089 TAA 1091

## RESULT 4

AAH16282

ID AAH16282 standard; cDNA; 2477 BP.

AC AAH16282;

DT 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:15148.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs

PS Claim 8; SEQ ID 15148; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 2477 BP; 738 A; 535 C; 524 G; 680 T; 0 other;

Query Match 100.0%; Score 1023; DB 22; Length 2477;  
Best Local Similarity 100.0%; Pred. No. 3.8e-310;  
Matches 1023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGTCAGTATGTCATCTCTTGTCTGATGTTGAAAGCTCATATCAACATTCGA 60  
DB 69 ATGTGTCAGTATGTCATCTCTTGTCTGATGTTGAAAGCTCATATCAACATTCGA 128  
QY 61 GATTCATATCCACCTCACACAGTGTCTAACAAGTCCATTTGAATCAACTAACAGGGCG 120  
DB 129 GATTCATATCCACCTCACACAGTGTCTAACAAGTCCATTTGAATCAACTAACAGGGCG 188  
QY 121 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATCTGCTGCT 180  
DB 189 AATCTTTTCTTCAAAATGTGAATCTTCCAGAAACAGGATCTTTTAAGATCTGCTGCT 248  
QY 181 CTCATATCCGTCAGAGTGTGTTCTCTGATGCTTTAGAAAGAACCCGAAAGCTGTGTT 240  
DB 249 CTCATATCCGTCAGAGTGTGTTCTCTGATGCTTTAGAAAGAACCCGAAAGCTGTGTT 308  
QY 241 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGAAAT 300  
DB 309 ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATTTGAAGAAAT 368  
QY 301 CCTGCTTATATGTGTCGCCAGACAGCTCCAGACTGTAAATAACTTTGCAATACAAGCC 360  
DB 369 CCTGCTTATATGTGTCGCCAGACAGCTCCAGACTGTAAATAACTTTGCAATACAAGCC 428  
QY 361 TACGAGCGTCAATTTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 420  
DB 429 TACGAGCGTCAATTTGTTACTGTGAACCTAGTGTAGTCCAGAGAAATGTTGCAAAA 488  
QY 421 AGAGTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCGCTGCAGTGATA 480  
DB 489 AGAGTTACAGAGAAACAGAGGCATCATGTTACATCCCAACAGGAGCGCTGCAGTGATA 548  
QY 481 GCTGCACAAGGGACAAATTTGCCCTGCAAGTGTGTAACACAGGTTCCCTTTGCTGGATCCTG 540  
DB 549 GCTGCACAAGGGACAAATTTGCCCTGCAAGTGTGTAACACAGGTTCCCTTTGCTGGATCCTG 608  
QY 541 GTGGTACCTGTAGGTGGAGGAGGAATGCTTGTGGAATAGCAATTTACAGTTAAAGCTCTG 600  
DB 609 GTGGTACCTGTAGGTGGAGGAGGAATGCTTGTGGAATAGCAATTTACAGTTAAAGCTCTG 668  
QY 601 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATGAGATGACTGTACAGTCC 660  
DB 669 AAACCTAGTGTGAAGGTATATGCTGCTGAACCCCTCAATGAGATGACTGTACAGTCC 728  
QY 661 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 720  
DB 729 AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 788

QY 721 AAATCCAGCATTTGGCTTGAAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATATCTTC 780  
DB 789 AAATCCAGCATTTGGCTTGAAACACCTGGCCCTATTATCAGGAGCCTTGTGGATGATATCTTC 848  
QY 781 ACTGTCCACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAACCTA 840  
DB 849 ACTGTCCACAGAGGATGAATTAAGTGTGCAACCCAGCTGGTGTGGAGAGGATGAACCTA 908  
QY 841 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 900  
DB 909 CTCATTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT 968  
QY 901 GTTTCCTCCAGAGTAAAGAACATTTGTAATTTGCTCAGTGGTGGAAATGTAGACTTAACC 960  
DB 969 GTTTCCTCCAGAGTAAAGAACATTTGTAATTTGCTCAGTGGTGGAAATGTAGACTTAACC 1028  
QY 961 TCCTCCATAAAGTGGTGAAGCAGCTGAAAGCCAGCTTCTATCAGTCTCTTCTGTT 1020  
DB 1029 TCCTCCATAAAGTGGTGAAGCAGCTGAAAGCCAGCTTCTATCAGTCTCTTCTGTT 1088  
QY 1021 TAA 1023  
DB 1089 TAA 1091  
RESULT 5  
AAI70575  
ID AAI70575 standard; cDNA; 1336 BP.  
XX AAI70575;  
AC AAI70575;  
XX 21-JAN-2002 (first entry)  
XX Human serine racemase cDNA.  
XX Serine racemase; human; D-serine; regulation;  
KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
KW nootropic; neuroprotective; cerebrotective; antiparkinsonian;  
KW analgesic; diagnosis; gene therapy; screening; ss.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH 52..1074  
FT CDS /\*tag= a  
ET  
XX  
PN WO200173077-A2.  
XX  
PD 04-OCT-2001.  
XX 30-MAR-2001; 2001WO-EP03668.  
XX 31-MAR-2000; 2000US-193748P.  
PR 03-APR-2000; 2000US-194249P.  
XX (FARB ) BAYER AG.  
PA  
XX Ramakrishnan S;  
PI  
XX WPI; 2001-648444/74.  
DR P-PSDB; AAM50262.  
XX Polynucleotide encoding serine racemase enzyme and the enzyme useful  
PT for screening reagents regulating the activity of the enzyme in a  
PT neuron disease caused by over- or under-activation of glutamate  
PT N-methyl-D-aspartate  
XX  
PS Claim 1; Fig 1; 66pp; English.  
XX  
CC The present sequence is that of cDNA encoding human serine racemase  
CC (see AAM50262). The polynucleotide can be used in the production



Query Match	99.8%;	Score 1021.4;	DB 22;	Length 2674;
Best Local Similarity	99.9%;	Pred. No. 1.3e-309;		
Matches 1022;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	ATGTGTGCTCAGTATTGTCATCTCTTTTGGCTGATGTTGAAAAAGCTCATATCAACATTCGA	60	
DB	262	ATGTGTGCTCAGTATTGTCATCTCTTTTGGCTGATGTTGAAAAAGCTCATATCAACATTCGA	321	
QY	61	GATTCTATCCACCTCACACCAGTGCTAACAAAGCTCCATTTTGAATCAACTTAACAGGCGC	120	
DB	322	GATTCTATCCACCTCACACCAGTGCTAACAAAGCTCCATTTTGAATCAACTTAACAGGCGC	381	
QY	121	AATCTTTTCTTCAAAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT	180	
DB	382	AATCTTTTCTTCAAAATGTGAAGTCTTCCAGAAACAGGATCTTTTAAAGATTCGTGGTCT	441	
QY	181	CTCAATGCCGTGAGAAGCTTGGTTCCTGATGCTTTAGAAAAGAAAGCCGAAAGCTGTGTT	240	
DB	442	CTCAATGCCGTGAGAAGCTTGGTTCCTGATGCTTTAGAAAAGAAAGCCGAAAGCTGTGTT	501	
QY	241	ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATGGGAAGGAATT	300	
DB	502	ACTCACAGCAGTGGAAACCATGGCCAGGCTCTCACCTATGCTGCCAAATGGGAAGGAATT	561	
QY	301	CCTGCTTATATTGTGTGGTCCCAAGACAGCTCCAGAGCTGTA AAAAAGCTGCAATACAGCC	360	
DB	562	CCTGCTTATATTGTGTGGTCCCAAGACAGCTCCAGAGCTGTA AAAAAGCTGCAATACAGCC	621	
QY	361	TACGGAGCGTCAATTGTTACTGTGAACCTAGTAGTGAGTCCAGAGAAAATGTTGCCAAA	420	
DB	622	TACGGAGCGTCAATTGTTACTGTGAACCTAGTAGTGAGTCCAGAGAAAATGTTGCCAAA	681	
QY	421	AGAGTTACAGAGAAACAGAAAGCATCATGGTATACATCCCAACAGGAGCGCTGCAGTGA	480	
DB	682	AGAGTTACAGAGAAACAGAAAGCATCATGGTATACATCCCAACAGGAGCGCTGCAGTGA	741	
QY	481	GCTGGACAAGGGACAATTGCCCTGGGAAGTGCTGAACACAGGTTCTTTGGTGGATGCAC	540	
DB	742	GCTGGACAAGGGACAATTGCCCTGGGAAGTGCTGAACACAGGTTCTTTGGTGGATGCAC	801	
QY	541	GTGGTACCTGTAGTGGAGGAGGAATGCTGTGCTGGAATAGCAATACAGTTAAGGCTCTG	600	
DB	802	GTGGTACCTGTAGTGGAGGAGGAATGCTGTGCTGGAATAGCAATACAGTTAAGGCTCTG	861	
QY	601	AAACCTTAGTGTGAAGGTATATGCTGTCTGAACCTCTCAAATGCAGATGACTGCTACCAG	660	
DB	862	AAACCTTAGTGTGAAGGTATATGCTGTCTGAACCTCTCAAATGCAGATGACTGCTACCAG	921	
QY	661	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC	720	
DB	922	AAGCTGAAGGGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGTTGTC	981	
QY	721	AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGACCTTGTGGATGATATCTTC	780	
DB	982	AAATCCAGCATTTGGCTTGAACACCTGGCTTATATCAGGAGACCTTGTGGATGATATCTTC	1041	
QY	781	ACTGTCCAGAGGATGAATTAAGTGTGAACCCAGCTGGTGTGGGAGAGGATGAACACTA	840	
DB	1042	ACTGTCCAGAGGATGAATTAAGTGTGAACCCAGCTGGTGTGGGAGAGGATGAACACTA	1101	
QY	841	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT	900	
DB	1102	CTCATTTGAACCTACAGCTGGTGTGGAGTGGCTGTGCTGTCTCAACATTTTCAAACT	1161	
QY	901	GTTTCCCCAGAGTAAGAACATTTGTATTGTCTCAGTGGTGGAAATGTAGACTTTAACC	960	
DB	1162	GTTTCCCTCAGAAGTAAGAACATTTGTATTGTCTCAGTGGTGGAAATGTAGACTTTAACC	1221	
QY	961	TCTCTCATAACTTTGGGTGAAGCAGGCTGAAAAGCCAGCTTCTTATCAGTCTGTTCTGTT	1020	
DB	1222	TCTCTCATAACTTTGGGTGAAGCAGGCTGAAAAGCCAGCTTCTTATCAGTCTGTTCTGTT	1281	
QY	1021	TAA	1023	

Db	1282 TAA 1284	
RESULT 7		
AAA59300		
ID	AAA59300 standard; DNA; 1672 BP.	
XX	AAA59300;	
XX		
XX	07-NOV-2000 (first entry)	
XX		
DE	DNA-encoding a murine serine racemase polypeptide.	
XX		
KW	Serine racemase; N-methyl-D-aspartate receptor; neural death;	
KW	neural dysfunction; NMDA receptor; Parkinson's disease;	
KW	Huntington's disease; motor neurone disease; Alzheimer's disease; ss.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	219..1238
FT		/*tag= a
FT		/product= "serine racemase"
XX		
FN	W0200043526-Al.	
XX		
PD	27-JUL-2000.	
XX		
PF	18-JAN-2000; 2000WO-US00938.	
XX		
PR	19-JAN-1999; 99US-0116333.	
PR	21-JUL-1999; 99US-0144839.	
PR	28-JUL-1999; 99US-0145953.	
XX		
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.	
XX		
PI	Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;	
PI	Ferris CD;	
XX		
DR	WPI; 2000-482915/42.	
DR	P-PSDB; AAB07731.	
XX		
PT	Mammalian serine racemase preparations, used to identify modulators	
PT	which can be used to treat diseases associated with	
PT	N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease	
PT		
XX		
PS	Disclosure; Fig 7A-B; 54pp; English.	
XX		
CC	The present sequence encodes a mammalian serine racemase, which has	
CC	a specific activity of at least 0.003 micromole L-serine/mg/hour.	
CC	The enzyme catalyses the direct racemisation of L-serine to D-serine.	
CC	D-serine appears to be an endogenous ligand of N-methyl-D-aspartate	
CC	(NMDA) receptors. The mammalian serine racemases can be used to identify	
CC	modulators, which can be used in the treatment of acute or chronic	
CC	neural death or dysfunction mediated by overactivation of N-methyl-D-	
CC	aspartate (NMDA) receptors. Overactivation of the receptors is	
CC	associated with Parkinson's disease, Huntington's disease, motor neurone	
CC	disease and Alzheimer's disease.	
XX		
SQ	Sequence 1672 BP; 491 A; 394 C; 379 G; 408 T; 0 other;	
	Query Match 79.8%; Score 816.6; DB 21; Length 1672;	
	Best Local Similarity 88.1%; Pred. No. 1.9e-245;	
	Matches 901; Conservative 0; Mismatches 119; Indels 3; Gaps 1	
Qy	1 ATGTGTCGTCAGTATTGTCATCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60	
Db	219 ATGTGTCGTCAGTACTGTCATCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 270	
Qy	61 GATTCTATCCACTACACAGTCGTGCTGCTGATTTTGAATCAACTAACAGGGCG 120	



Db 279 GACTCTATCCACCTCACCCAGTGGCTAACAGAGCTCCATTTTGAATCAATAGCAGGGCGC 338  
Qy 121 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGAGTCTTTTAAAGATTGGTGGTCT 180  
Db 339 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGAGTCTTTTAAAGATTGGTGGTGGC 398  
Qy 181 CTCATGCGCTGACAGCTTGGTCTTCCATGATGCTTTAGAAAGGAGCGGAAAGCTGTGTT 240  
Db 399 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCGCAGAGAGAGAGCCCAAGCCGTAGTT 458  
Qy 241 ACTCACAGCAGTGAAGAACCATGCGCAGGCTCTCACTATGCTGCCAAATTCGAGGAATT 300  
Db 459 ACTCACAGCAGCGGAACCATGCGCAGGCTCTCACTATGCTGCCAAATTCGAGGAATT 518  
Qy 301 CCTGCTTATATGTGGTGGCCAGACAGCTCCAGACTGTAAAACTTGAATAACAAGCC 360  
Db 519 CTTGCTTACATTTGGTGTCCCAACAGCTCCCAACTGCAAGAACTGGCAATCCAAGCC 578  
Qy 361 TACGAGGCTCAATTTGATATGTAACCTAGTAGTCCAGTCCAGAGAAAATGTTCGAAA 420  
Db 579 TATGGAGCATCGATAGTATAGTGTGACCAAGTGACGAGTCCAGAGAAAAGGTCACCTCA 638  
Qy 421 AGAGTTACAGAGAAACAGAGGAGCATGATGATACATCCCAACAGGAGCTGCAAGTATA 480  
Db 639 AGAATTTATGAAAGAAACAGAGGAGCATGTTGCTCATCCCAACAGGAGGCTGCAAGTATA 698  
Qy 481 GCTGACAGGAGCAATTTGCCCTGGAAGTGTGTAACCAAGGTTCTTTGGTGGATGCACTG 540  
Db 699 GCTGGACAGGAACAAATTTGCCCTGGAAGTGTGTAACCAAGGTTCTTTGGTGGATGCACTG 758  
Qy 541 GTGTAAGTGTGAGGAGGAGGAGTCTTGGTGGAAATAGCAATACAGTTAAAGGCTCTG 600  
Db 759 GTGTAAGTGTGAGGAGGAGGAGTCTTGGTGGAAATAGCAATACAGTTAAAGGCTCTG 818  
Qy 601 AAACCTAGTGTGAGGAGTATGCTGTAACCTCAACCTCAATGCAAGTACTGCTACCAGTCC 660  
Db 819 AAACCTAGTGTGAGGAGTATGCTGTAACCTCAACCTCAATGCAAGTACTGCTACCAGTCT 878  
Qy 661 AAGCTGAAGGGGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 720  
Db 879 AAACCTGAAGGAGAACTGATGCCCAATCTTTATCTCTCCAGAAACCATAGCAGATGCTGTC 938  
Qy 721 AAATCCAGCATTTGGCTGACACCTGGCCTTATATCAGGAGACCTTGTGGATGATATCTTC 780  
Db 939 AAATCCAGCATTTGGCTGAACTGCTGGCTTATTAAGAGACCTTGTGGATGATGCTTC 998  
Qy 781 ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTA 840  
Db 999 ACTGTACAGAGGATGAAATTAAGTGTGCAACCCAGCTGTGTGGAGAGGATGAACCTG 1058  
Qy 841 CTCATTGAACCTACAGTGTGTGGAGTGGCTGTGTGCTGTCTCAACATTTTCAAACT 900  
Db 1059 CTCATTGAACCTACAGTGTGTGGAGTGGCTGTGTGCTGTCTCAACATTTTCAAACT 1118  
Qy 901 GTTTCCCAAGAACTAAGACATTTGATTTGCTCTCAGTGGTGAATAGTACTTAACC 960  
Db 1119 GTCTCTCCAGAACTAAGACATTTGATTTGCTCTCAGTGGTGAATAGTACTTAACC 1176  
Qy 961 TCTCCATAACTTTGGGTGAAGCAGCTGAAAGCCAGCTTCTTATCAGTGTGTTCTGTT 1020  
Db 1177 -CTCCCTGNACTGGGTGGGCGAGGCTGAACGCCAGCTCTCTTACCAGAGGGTTTCTGTT 1235  
Qy 1021 TAA 1023  
Db 1236 TAA 1238

## RESULT 8

AAA59294

ID AAA59294 standard; DNA; 1018 BP.

XX

AC AAA59294;

XX

DT 07-NOV-2000 (first entry)  
XX DNA encoding a murine serine racemase polypeptide.  
DE  
XX Serine racemase; N-methyl-D-aspartate receptor; neural death;  
KW neural dysfunction; NMDA receptor; Parkinson's disease;  
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.  
XX  
OS Mus musculus.  
XX Key Location/Qualifiers  
FH 1..1018  
CDS /\*tag= a  
FT /transl\_except= (pos: 1012, aa: Ser)  
FT /product= "serine racemase".  
XX  
PN WO200043526-A1.  
XX  
XX 27-JUL-2000.  
XX  
XX 18-JAN-2000; 2000WO-US00938.  
XX  
XX 19-JAN-1999; 99US-0116333.  
PR 21-JUL-1999; 99US-0144839.  
PR 28-JUL-1999; 99US-0145953.  
XX  
XX (UOJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
XX  
XX Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;  
PI Ferris CD;  
XX  
XX WPI: 2000-482915/42.  
DR P-PSDB; AAB07731.  
XX  
XX Mammalian serine racemase preparations, used to identify modulators  
PT which can be used to treat diseases associated with  
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease  
PT  
XX  
XX Claim 16; Page 45-46; 54pp; English.  
XX  
XX The present sequence encodes a mammalian serine racemase, which has  
CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
CC modulators, which can be used in the treatment of acute or chronic  
CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
CC aspartate (NMDA) receptors. Overactivation of the receptors is  
CC associated with Parkinson's disease, Huntington's disease, motor neurone  
CC disease and Alzheimer's disease.  
XX  
SQ Sequence 1018 BP; 290 A; 250 G; 243 G; 235 T; 0 other;

Query Match 79.0%; Score 807.8; DB 21; Length 1018;

Best Local Similarity 87.7%; Pred. No. 8.4e-243;

Matches 894; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

Qy 1 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
Db 1 ATGTGTGCTCAGTATTGTCATCTCTTTGCTGATGTTGAAAAGCTCATATCAACATTCGA 60  
Qy 61 GATTTCTATCCCTCACCCAGTGTCAACAGCTCCATTTGTAATCAACTAACAGGGCGC 120  
Db 61 GACTCTATCCCTCACCCAGTGTCAACAGCTCCATTTGTAATCAACTAACAGGGCGC 120  
Qy 121 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGAGTCTTTTAAAGATTGGTGGTCT 180  
Db 121 AATCTTTTCTTCAAAATGAACTCTTCCAGAAAACAGAGTCTTTTAAAGATTGGTGGTCT 180  
Qy 181 CTCATGCGCTGACAGCTTGGTCTGATGCTTTAGAAAGGAGCGGAAAGCTGTGTT 240  
Db 181 CTTAATGCCATCAGAGGCTTAATCTCTGACAGCGCAGAGAGAGCCCAAGCCGTAGTT 240



241	QY	ACTCACAGCATGTGGAAACCATGGCCAGGCTCTACCTATGCTGCCAAATTTGGAAGGAATTT	300
241	Db		
241	QY	ACTCACAGCGCGGAAACCATGGCCAGGCTCTACCTATGCTCTAACTGGGAAGGAATTT	300
301	QY	CCTGCTTATATTGTGGTCCGCCAGACAGCTCCAGACTGTAAAAAACCTTGCATATACAAGCC	360
301	Db		
301	QY	CCTGCTTACATTGTGGTTCCTCCCAACAGCTCCCACTGCAAGAAACTGGCAATGCCAGCC	360
361	QY	TACGGAGCGCTCAATTGTATCTGTGCAACCTAGTGATGAGTCCAGAGAAAATTTGTGCAAAA	420
361	Db		
361	QY	TATGGAGCATCGATAGTATATGCTGACCCAAGTGACGAGTCCAGAGAAAAGGTCTACTCAA	420
421	QY	AGAGTTACAGARAAGAAACAGAGGAGCATCATGGTACATCCCAACAGAGAGCTCGAGTGATA	480
421	Db		
421	QY	AGAAATTATGCAAGAAACAGAGGAGCATCTTGTTCATCCCAACAGAGAGCTCGCAGTGATA	480
481	QY	GCTGCACAGGAGCAAAATTCGCCCTGGAAAGTGCTGAACAGGTTCTCTTTGGTGGATGCACGTG	540
481	Db		
481	QY	GCTGCACAGGAGCAAAATTCGCCCTGGAAAGTGCTGAACAGGTTCTCTTTGGTGGATGCACGTG	540
541	QY	GTGGTACTCTGTAGTGGAGGAGGAATGCTTGCTGGAATAGCAATACAGTTAAGGCTCTG	600
541	Db		
541	QY	GTGGTACCAGTAGGAGGAGGAGGAATGTTGCTGGAATAGCCATTACAAATTAAGGCGCTG	600
601	QY	AAACCTAGTGTCAAGGTATATGCTGCTGAACCCCTCAATGCAGATGACCTGCTACCAAGTCC	660
601	Db		
601	QY	AAACCTAGTGTGAAGGTATAGCTGCTGAGCCCTCGAATGCAGATGACCTGCTACCAAGTCT	660
661	QY	AAAGCTGAAGGGGAAACTGATGCCAAATCTTTATCTCCAGAAACCATAGCAGATGGTGTC	720
661	Db		
661	QY	AAACTGAAGGAGAACTGACCCCACTTCATCTCCAGAAACCATAGCAGATGGTGTC	720
721	QY	AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTTGCGATGATATCTTC	780
721	Db		
721	QY	AAATCCAGCATTTGGCTTGAACACCTGGCCCTATTATCAGGACCTTTGCGATGATATCTTC	780
781	QY	ACTGTCCAGAGGATGAATTAAGCTGTGCACACCCAGCTGGTGGGAGGAGTGAACACTA	840
781	Db		
781	QY	ACTGTCCAGAGGATGAATTAAGCTGTGCACACCCAGCTGGTGGGAGGAGTGAACACTA	840
841	QY	CTCATTTGAACCTACAGCTGGTGTGGAGTGCGTCTGTCTGTCTCAACATTTTCAAACT	900
841	Db		
841	QY	CTCATTTGAACCTACAGCTGGTGTGGAGTGCGTCTGTCTGTCTCAACATTTTCAAACT	900
901	QY	GTTTCCCAAGAGTAAGAAACATTTGATTTGCTCAGTGGTGGAAATGTAGACTTTAAAC	960
901	Db		
901	QY	GTCTCTCCAGAGTAAGAAAGCTGTGCAATTTGACTCAGTGGGGGGAATGTAGACCTTAA	958
961	QY	TCCTCCATAACTTTGGGTGCAAGCAGGCTTCAAAAGGCCAGCTTCTTTATCAGTCTGTTCTCTG	1019
959	Db	---CCTCCCTGAACCTGGGTGGGCGAGGCTGAACGGCCAGCTCCTTTACAGAGCGTCTGTTT	1016
RESULT 9			
AAH06600			
ID	AAH06600 standard; cDNA; 848 bp.		
AC	AAH06600;		
XX	26-JUN-2001 (first entry)		
DT	Human cDNA clone (5'-primer) SEQ ID NO:3435.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	Homo sapiens.		
OS	EP1074617-A2.		
PN	XX		
XX	07-FEB-2001.		
PD	28-JUL-2000; 2000EP-0116126.		
XX			



```

RESULT 11
AAA59295
ID AAA59295 standard; DNA; 608 BP.
XX
AC AAA59295;
XX
DT 07-NOV-2000 (first entry)
XX
DE N-terminal sequence of human serine racemase DNA.
XX
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;
KW neural dysfunction; NMDA receptor; parkinson's disease;
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.
XX
OS Homo sapiens.
XX
PN WO200043526-A1.
XX
PD 27-JUL-2000.
XX
PF 18-JAN-2000; 2000WO-US00938.
XX
PR 19-JAN-1999; 99US-0116333.
PR 21-JUL-1999; 99US-0144839.
PR 28-JUL-1999; 99US-0145953.
XX
PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;
PI Ferris CD;
XX
DR WPI; 2000-482915/42.
XX
PT Mammalian serine racemase preparations, used to identify modulators
PT which can be used to treat diseases associated with
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease
XX
PS Claim 17; Page 26; 54pp; English.
XX
CC The present sequence represents a fragment of a mammalian serine
CC racemase gene. The racemase polypeptide has a specific activity of at
CC least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the
CC direct racemisation of L-serine to D-serine. D-serine appears to be
CC an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The
CC mammalian serine racemases can be used to identify modulators, which
CC can be used in the treatment of acute or chronic neural death or
CC dysfunction mediated by overactivation of N-methyl-D-aspartate (NMDA)
CC receptors. Overactivation of the receptors is associated with
CC Parkinson's disease, Huntington's disease, motor neurone disease and
CC Alzheimer's disease.
XX
SQ Sequence 608 BP; 174 A; 137 C; 148 G; 148 T; 1 other;

Query Match 52.3%; Score 535; DB 21; Length 608;
Best Local Similarity 98.4%; Pred. No. 2.8e-157;
Matches 571; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

QY 1 ATGTGTGTCAGTATTGCTATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 60
DB 27 ATGTGTGTCAGTATTGCTATCTCCCTTTGCTGATGTTGAAAAAGCTCATATCAACATTCGA 86
QY 61 GATTCTATCCACTCACACGAGTGTAAACAAGCTCCATTTTGAATCAACTACAGGGCCG 120
DB 87 GATTCTATCCACTCACACGAGTGTAAACAAGCTCCATTTTGAATCAACTACAGGGCCG 146
QY 121 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAAAGAGGATCTTTTAAGATTGCTGTGCT 180
DB 147 AATCTTTTCTTCAATGTGAAGTCTTCCAGAAAAAGAGGATCTTTTAAGATTGCTGTGCT 206
QY 181 CTCATGCGCGTCAGAGCTTGTTCTCTGATGCTTTTAGAAAAAGGACCGCAAGCTGTGTT 240

```

```

|||||
207 CTCATGCGCGTCAGAGCTTGTTCTCTGATGCTTTAGAAAAAGGACCGCAAGCTGTGTT 266
QY 241 ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAATT 300
DB |||||||
267 ACTCACAGCAGTGGAAACCATGCCAGGCTCTCACCTATGCTGCCAAATTTGGAGGAATT 326
QY 301 CCTGCTTATATTGCTGTCGCCCCACAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 360
DB |||||||
327 CCTGCTTATATTGCTGTCGCCCCACAGACAGCTCCAGACTGTAAAAAATTTGCAATACAGCC 386
QY 361 TAGGGAGCGTCAATTCTATCTATGTAACCTAGTGATG-AGTCCAGAGAAAAATTTGCCAAA 419
DB |||||||
387 TAGGGAGCGTCAATTCTATCTATGTAACCTAGTGATG-AGTCCAGAGAAAAATTTGCCAAA 446
QY 420 AA-GAGTTTACAGAAAGACAGAGGATCATGCTATCATCCCAACAGGACCTTGCAGTGA 478
DB || |||||||
447 AAGGAGTTTACAGAAAGACAGAGGATCATGCTATCATCCCAACAGGACCTTGCAGTGA 506
QY 479 TAGCTGACACAAGGACCAATTGCCCTGGAAGTGTGAACAGAGTTCCCTTTGGTGGATGCAC 538
DB |||||||
507 TAGCTGACACAAGGACCAATTGCCCTGGAAGTGTGAACAGAGTTCCCTTTGGTGGATGCAC 566
QY 539 TGGTGTACTCTGATGTTGG-AGGAGGAATGCTTGTCTGGAA 577
DB |||||
567 TGGTGTGNCCTCTGATGTTGGAAAGGAGGAATGCTTGTCTGGCGGA 606

RESULT 12
AAA59296
ID AAA59296 standard; DNA; 509 BP.
XX
AC AAA59296;
XX
DT 07-NOV-2000 (first entry)
XX
DE C-terminal sequence of human serine racemase DNA.
XX
KW Serine racemase; N-methyl-D-aspartate receptor; neural death;
KW neural dysfunction; NMDA receptor; parkinson's disease;
KW Huntington's disease; motor neurone disease; Alzheimer's disease; ss.
XX
OS Homo sapiens.
XX
PN WO200043526-A1.
XX
PD 27-JUL-2000.
XX
PF 18-JAN-2000; 2000WO-US00938.
XX
PR 19-JAN-1999; 99US-0116333.
PR 21-JUL-1999; 99US-0144839.
PR 28-JUL-1999; 99US-0145953.
XX
PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;
PI Ferris CD;
XX
DR WPI; 2000-482915/42.
XX
PT Mammalian serine racemase preparations, used to identify modulators
PT which can be used to treat diseases associated with
PT N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease
XX
PS Claim 18; Page 27; 54pp; English.
XX
CC The present sequence represents a fragment of a mammalian serine
CC racemase gene. The racemase polypeptide has a specific activity of at
CC least 0.003 micromole L-serine/mg/hour. The enzyme catalyses the
CC direct racemisation of L-serine to D-serine. D-serine appears to be
CC an endogenous ligand of N-methyl-D-aspartate (NMDA) receptors. The

```

CC mammalian serine racemases can be used to identify modulators, which  
CC can be used in the treatment of acute or chronic neural death or  
CC dysfunction mediated by overactivation of N-methyl-D-aspartate (NMDA)  
CC receptors. Overactivation of the receptors is associated with  
CC Parkinson's disease, Huntington's disease, motor neuron disease and  
CC Alzheimer's disease.  
XX  
SQ Sequence 509 BP; 139 A; 90 C; 115 G; 164 T; 1 other;

Query Match 32.98; Score 337; DB 21; Length 509;  
Best Local Similarity 99.78; Pred. No. 3.7e-95;  
Matches 348; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
Qy 676 CTGATGCCCAATCTTTATCTCCAGAACCATAGCAGATGGTGCAAAATCCAGCATGGC 735  
Db 1 CTGATGCCCAATCTTTATCTCCAGAACCATAGCAGATGGTGCAAAATCCAGCATGGC 60  
  
Qy 736 TTGAA-CACCTGGCCTATTATCAGGGACCTTTGGGAGGATGATCTTCACTGTCACAGGA 794  
Db 61 TTGAANACCTGGCCTATTATCAGGGACCTTTGGGAGGATGATCTTCACTGTCACAGGA 120  
  
Qy 795 TGAATTAAGTGTCAACCCAGCTGGTGCGGAGGATGAACTACTGTAACCTAC 854  
Db 121 TGAATTAAGTGTCAACCCAGCTGGTGCGGAGGATGAACTACTGTAACCTAC 180  
  
Qy 855 AGCTGGTGTGGAGTGCTGTGCTGTCTCAACATTTTCAAACTGTTTCCCCAGAAAT 914  
Db 181 AGCTGGTGTGGAGTGCTGTGCTGTCTCAACATTTTCAAACTGTTTCCCCAGAAAT 240  
  
Qy 915 AAGAACAATTTGATTTGCTGCTAGTGTGGAATGTAGACTTAACCTCCCTCCATACTTG 974  
Db 241 AAGAACAATTTGATTTGCTGCTAGTGTGGAATGTAGACTTAACCTCCCTCCATACTTG 300  
  
Qy 975 GGTGAACGAGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 1023  
Db 301 GGTGAACGAGCTGAAGGCCAGCTTCTTATCAGTCTGTTTCTGTTTAA 349

RESULT 13  
ABAI5709/C  
ID ABAI5709 standard; DNA; 20892 BP.  
XX ABAI5709;  
AC ABAI5709;  
DT 23-JAN-2002 (first entry)  
XX Human nervous system related polynucleotide SEQ ID NO 8040.  
DE  
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX Homo sapiens.  
XX WO200159063-A2.  
XX 16-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-02411787.  
PR 20-OCT-2000; 2000US-02411808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0242221.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
PS Disclosure; SEQ ID NO 8040; 1701pp + Sequence Listing; English.  
XX

XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins

CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (antagonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 20892 BP; 5768 A; 4125 C; 5164 G; 5835 T; 0 other:

Query Match 21.6%; Score 221; DB 22; Length 20892;

Best Local Similarity 100.0%; Pred. No. 8e-58;

Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGTG 862  
|||||  
Db 12794 AGTGTGCAACCCAGCTGGTGGGAGAGGATGAACCTACTTGAACCTACAGCTGGTG 12735  
QY 863 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTCCCCAGAGTAAGAACA 922  
|||||  
Db 12734 TTGGAGTGGCTGCTGCTCTCAACATTTTCAAACTGTTCCCCAGAGTAAGAACA 12675  
QY 923 TTTGTATTGTCAGTGGTGGAAATCTAGACTTAACCTCCTCATAAATTTGGTGAAGC 982  
|||||  
Db 12674 TTTGTATTGTCAGTGGTGGAAATCTAGACTTAACCTCCTCATAAATTTGGTGAAGC 12615  
QY 983 AGGCTGAAGGCCAGCTTCTTATCAGTCTCTTTCTGTTTAA 1023  
|||||  
Db 12614 AGGCTGAAGGCCAGCTTCTTATCAGTCTCTTTCTGTTTAA 12574

#### RESULT 14

AAC10869

ID AAC10869 standard; cDNA; 861 BP.

XX AAC10869;

XX 06-OCT-2000 (first entry)

DT Human secreted protein 5' EST, SEQ ID NO: 14944.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.

OS EP1033401-A2.

PN 06-SEP-2000.

PD 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 14944; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively



[illegible]

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 24, 2003, 06:21:15 ; Search time 1731 Seconds  
(without alignments)  
867.232 Million cell updates/sec

Title: US-09-889-609B-9  
Perfect score: 1023  
Sequence: 1 atgtgtctcagttatgc.....atcagtcgtttctgtttaa 1023

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues 2085038

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.4	14.0	31096	7 US-08-781-986A-59	Sequence 59, Appl
2	89.4	8.7	1830121	9 US-10-329-960-1	Sequence 1, Appl
3	77.4	7.6	105184	9 US-09-847-513A-1	Sequence 1, Appl
4	70.6	6.9	954	10 US-09-974-300-4749	Sequence 1091, Ap
5	69.6	6.8	930	9 US-09-738-626-1091	Sequence 1, Appl
6	69.6	6.8	3309400	9 US-09-738-626-1091	Sequence 1, Appl
7	57	5.6	1545	9 US-09-942-891-1	Sequence 5, Appl
8	57	5.6	1545	9 US-09-942-891-5	Sequence 7, Appl
9	57	5.6	1545	9 US-09-942-891-7	Sequence 8, Appl
10	57	5.6	1545	9 US-09-942-891-8	Sequence 264, App
11	49.4	4.8	1251	10 US-09-974-300-264	Sequence 1, Appl
12	48.6	4.8	536165	9 US-09-939-964-1	Sequence 102, App
13	48.4	4.7	926	10 US-09-974-300-4809	Sequence 207, App
14	46.2	4.5	15249	7 US-08-781-986A-102	Sequence 49, Appl
15	46	4.5	2698	10 US-09-939-980-207	Sequence 47, Appl
16	38.6	3.8	1655	9 US-10-234-432-49	Sequence 50, Appl
17	38.6	3.8	1855	9 US-10-234-432-47	Sequence 46, Appl
18	38.6	3.8	2656	9 US-10-234-432-50	
19	38.6	3.8	2675	9 US-10-234-432-46	

Sequence 197, Appl  
Sequence 1928, Ap  
Sequence 18, Appl  
Sequence 21, Appl  
Sequence 3, Appl  
Sequence 51, Appl  
Sequence 271, Appl  
Sequence 2742, Ap  
Sequence 3356, Ap  
Sequence 11109, A  
Sequence 59, Appl  
Sequence 1627, Ap  
Sequence 11, Appl  
Sequence 28, Appl  
Sequence 455, Appl  
Sequence 212, Appl  
Sequence 210, Appl  
Sequence 9, Appl  
Sequence 5804, Ap  
Sequence 5803, Ap  
Sequence 32623, A  
Sequence 75, Appl  
Sequence 1101, Ap  
Sequence 5906, Ap  
Sequence 10, Appl  
Sequence 1, Appl

38.4 3.8 1379 9 US-10-098-841-197  
37.2 3.6 1827 10 US-09-974-300-1928  
37.2 3.6 1833 10 US-09-921-823-18  
37.2 3.6 1833 10 US-09-921-823-21  
36.6 3.6 1200 10 US-09-921-823-3  
35.4 3.5 1034 9 US-10-234-432-51  
35 3.4 928 10 US-09-974-300-271  
35 3.4 2000 9 US-09-938-842A-2742  
35 3.4 2000 9 US-09-938-842A-3356  
34 4.27 10 US-09-960-352-11109  
34 3.4 915 9 US-10-169-048-59  
34.2 3.3 1359 9 US-09-938-842A-1627  
34.2 3.3 118951 9 US-10-161-572-11  
33.8 3.3 763 9 US-10-153-668-28  
33.8 3.3 790 9 US-10-153-668-455  
33.8 3.3 880 10 US-09-822-849A-212  
33.8 3.3 1232 9 US-10-098-841-210  
33.8 3.3 3257 12 US-10-029-654-9  
33.8 3.3 3773 9 US-09-764-891-5804  
33.8 3.3 3774 9 US-09-764-891-5803  
33.6 3.3 486 9 US-09-918-995-32623  
33.4 3.3 1274 9 US-09-983-802-75  
33.2 3.2 873 9 US-09-738-626-1101  
33.2 3.2 936 9 US-09-738-626-1101  
33.2 3.2 173808 12 US-10-003-806-10  
33.2 3.2 3309400 9 US-09-738-626-1

#### ALIGNMENTS

#### RESULT 1

US-08-781-986A-59

Sequence 59, Application US/08781986A

Publication No. US20030054436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob

REGISTRATION NUMBER: 30,446

REFERENCE/DOCKET NUMBER: PB248pp

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 31096 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-781-986A-59

Query Match		14.0%	Score 143.4;	DB 7;	Length 31096;
Best Local Similarity		48.6%	Pred. No. 2.4e-34;		
Matches		456;	Conservative	0;	Mismatches 472; Indels 10; Gaps 2;
Qy	25	TTTGCTGATGTTGAAGGCTCATATACAACTTCGAGATTCATCCACCTTCACACGAGTG	84		
Db	1251	TTAGGAGATATCGAAGGCTTAAGCAAGCATTTAAACCAATTTATTCGTCGAACACCTCTA	1310		
Qy	85	CTAACAGCTCCATTTTGATCAA---CTAACAGGCGCAATCTTTTCTTCAATGTGA	141		
Db	1311	ATTAATCAATGATTTTAGCGAAGATTAACATAAGGGAATGATTTCTTAAATTAGAA	1370		
Qy	142	CTCTCCGAAAAACAGGATCTTTTAAGATTCGFGTCTCTCAATGCGCTCAAGAGCTTG	201		
Db	1371	AATATGCAATTCACAGGATCTTTTAATTTAGAGCGCTAGCAAT-----NAAATTA	1423		
Qy	202	GTTCCTGATGCTTTAGAAAGGAGCGCAAGCTGTTGTTACTCACAGCAGTGGAACCAT	261		
Db	1424	ATCACTTAACAGATGAACAAAAAGAAAAGGCATTTATCGCAGCATCTGCTGGGGAACCAT	1483		
Qy	262	GGCAGGCTCTACCTATGTCGCAATTCGAAGGATTCCTGCTTATTTGTTGGCGCC	321		
Db	1484	GCACAGGTGCTCTTAACAGCTAAATTTATAGGCATTTGATGCAACGATTTGTAAGCCT	1543		
Qy	322	CAGACAGCTCCAGACTGTAAAAAATTGCAATACAAGCCTACGAGCGCTCAATTTGTATAC	381		
Db	1544	GAACAGCACCAAGCAAGCAACCAACCAAAAGGCTATGGGCAAAAGGTTATTTTA	1603		
Qy	382	TGTGAACCTAGTGATGAGTCCAGAGAAATGTTGCAAAAGATTTACAGAAGAAACAGAA	441		
Db	1604	AAAGGTAAAAAATTTAACGAACTAGACTTTATATGAAGAATTTAGCGAAAGAAATGGC	1663		
Qy	442	GGCATCATGTGTACATCCCAACAGGAGCGCTGCGAGTGATAGCTGCAAGGGAACATGGC	501		
Db	1664	ATCACAATGCTTCATCCATATGACATAGATTTGTAATGCGAGCGCAAGCAAAATTTGGT	1723		
Qy	502	CTGGAAGTGTGAACAGGTTCTTTTGGTGGATGCACTGGTGGTACCTGTAGTGAGGGA	561		
Db	1724	TTAGAAATTTTAGATGATATTTGGAATGTGAATACAGTCATCGTCCAGTTGGCGGTGA	1783		
Qy	562	GGAATGCTGCTGGAATACCAATTTACAGTTAAGCTCTGNAACCTAGTGTGANGGTATAT	621		
Db	1784	GGATTAAATTCAGGATTTGCCACCGCATTTAAATCAATTTAAACCTTCAATTCATATATC	1843		
Qy	622	GCTGCTGAACCCCTCAATGCAGATGACTGCTACCAAGTCCAAAGTGAAGGGGAACATGATG	681		
Db	1844	GGTGTTCATCTCAGATGTTTCATGGTATGGCTGAGTCTTTCTATAGAGAGATTTAACT	1903		
Qy	682	CCCAATCTTTATCCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTGAAC	741		
Db	1904	GAACATCGAGTGGATAGCAATAGCAGATGGTTGTGTATAAAGTTCTCTGTGTAACAA	1963		
Qy	742	ACCTGGCCTATTATCAGGACCTTTGGATGATATCTTCACTGTCACAGGATGAAAT	801		
Db	1964	ACATATGAAGTATTAACCATTTAGTAGATGATTTATTTCTTTGTTACTGAAGAAGAAAT	2023		
Qy	802	AAGTGTGCAACCCAGCTGTTGGGAGAGGATGAAACTACTCATTTCAACCTACAGCTGT	861		
Db	2024	GAACATGCTATGAAGATTTAATGCAGCGTGCACAAATTTACTGAAGGTGCAGCGCA	2083		
Qy	862	GTTGGAGTGGCTGCTGTGTCTCAACATTTTCAAACTGTTTCCCCAGAAGTAAAGAAC	921		
Db	2084	TTACCAACAGCTGCAATTTTAAGTGGAAAAATAAACAATAAATGGCTTGAAGATAAAAT	2143		
Qy	922	ATTTGATTTGCTCAGTGGTGAATGTAGACTTTAAC	959		
Db	2144	GTTGTTGCATAGTTTTCAGGCGGGAATGTTGACTTAAC	2181		

RESULT 2  
US-10-329-960-1  
; Sequence 1, Application US/10329960  
; Publication No. US20030099277A1

GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F  
; FILE REFERENCE: PB186PI  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (40810)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc\_feature

```
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Query Match      8.7%; Score 89.4; DB 9; Length 1830121;
Best Local Similarity 46.4%; Pred. No. 8.7e-16;
Matches 291; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

QY 253 GGAACCATGCCAGCCTCTCACCTATGCTGCCAAATTCGAAGCAATTCCTGCTTATATT 312
Db 794225 GGTAAACCATGCCAGCGGTGGCATTATCGCGCAACAATTCAGCTTTAAAGCATTAATT 794284

QY 313 GTGGTCCCCAGACAGCTCCAGACTGTAAAAAATTCGAATACAAAGCTACGGAGCGTCA 372
Db 794285 GTTATGCCACAAAACACCCCAAGCATTAAGTGGATGCATGCGTGTGGTGGTGAG 794344

QY 373 ATGTATACGTGAACCTAGTGTATGATGTCAGAGAAAATGTTCAAAAAGAGTTACAGAA 432
Db 794345 GTGTTGTCACGGTGTCTAATTTTCGATGAAGCCAAAAGCAATTCGAGCTTTCAAAA 794404

QY 433 GAACAGAGGCATCATGTTGATACATCCCAACGAGCGCTCAGTAGCTGACAGGG 492
Db 794405 GAAAAAACAATGACATTTATTCCACCATTCGATCATCCATTAGTAGTCGAGCAAGGC 794464
```

Qy	493	ACAAATTCGCCCTGGAAGTGGTGAACACAGGTTCTCTTTGGTGGATGCACGTGGTGTAACCTGTA	552
Db	794465	ACCTTAGCAGTGAAGTATGCTACAAACAGTGGCGGATTTGGGATATGTGTGTACAAGTT	794524
Qy	553	GGTGAGGAGGAATCCTTGCTGGAAATACGAATACAGTTAAGGCTCTGAAACCTAGTGTG	612
Db	794525	GGTGGTGGTGGCTTAGCTGCGGGGTGGCAATTTTGCTTAAGCAATTTATGCCGGAAAT	794584
Qy	613	AAGGTATATGCTGCTGAACCCCTCAAAATGCAGATGACTGCTACCAGTCCAAGCTGAAGGGG	672
Db	794585	AAATCATCGGTGTAGATCAAAAGGATTTCTCGTGCTTAAGCGGCTCTCGATAAAGC	794644
Qy	673	AAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTCAAATCCAGCAT	732
Db	794645	GAACCAACAGATTTAAACCATATTGGATTATTTGCCGATGGCGTTGCTGTAACACGCAT	794704
Qy	733	GCCTTGAACACCTGGCCCTATTATCAGGACCTTCGGATGATATCTTCACTGTGCACAG	792
Db	794705	GGCAGGAAACATTCGGTCTATGTACGCAATATCTTGATGATATGGTATTTGGTCGATAGT	794764
Qy	793	GATGAAATTAAGTGTGCAACCCAGCTGTGTGGGAGAGGATGAAACTTACTCATTTGAACCT	852
Db	794765	GACGAAGTATGCCAGCAATGAAGATTTGTTGAGAACGTTCTGCTGTGTGCAGAACCA	794824
Qy	853	ACAGCTGGTGTGGAGTGGCTGTGTG	879
Db	794825	TCAGTGCATATAGGTTTGGCTGGTTG	794851

```

RESULT 3
US-09-847-513A-1
; Sequence 1, Application US/09847513A
; Publication No. US20030104375A1
; GENERAL INFORMATION:
; APPLICANT: MBARI
; APPLICANT: DeLong, Edward
; APPLICANT: Beja, Oded
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin
; FILE REFERENCE: MBA-101
; CURRENT APPLICATION NUMBER: US/09/847,513A
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,602
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 105184
; TYPE: DNA
; ORGANISM: Naturally occurring gamma proteobacterium
; FEATURE:
; NAME/KEY: gene
; LOCATION: (50866)..(51615)
; OTHER INFORMATION: Proteorhodopsin gene sequence.
; NAME/KEY: misc.feature
; LOCATION: (1593)..(2807)
; OTHER INFORMATION: Predicted threonine dehydratase. Contains 'n' at position 2753
; PUBLICATION INFORMATION:
; AUTHORS: Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P.,
; AUTHORS: Jovanovich, S.B., Gates, C.M., Feldman, R.A., DeLong, E.F.
; TITLE: Bacterial rhodopsin: evidence for a new type of phototrophy in the sea
; JOURNAL: Science
; VOLUME: 289
; ISSUE: 5486
; PAGES: 1902-1906
; DATE: 2000-09-15
; DATABASE ACCESSION NUMBER: AF279106
; DATABASE ENTRY DATE: 2000-06-15
; RELEVANT RESIDUES: (50866)..(51615)
US-09-847-513A-1

```

Query Match 7.6%; Score 77.4; DB 9; Length 105184;  
Best Local Similarity 51.0%; Pred. No. 9.7e-13;  
Matches 183; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy	223	AAGCCGAAAGTGTTGTTACTACAGCAGGTGGAAACCATGGCCAGGCTCTCACCTATGCT	282
Db	1500	AAGAAGAGGGGGTTATTGCTGCATCAGCAGGAAATCATGCTCAAGGGGTAGCCAGTGCA	1559
Qy	283	GCCAAATTGGAAGGAATTCCTGCTTATATTGTCGTGCCCCAGACAGCTCCAGACTGTAAA	342
Db	1560	TGTAAAGAAATTAATAAATTAATGCTTGATAGTTATGCCAATAACAACCTCCAGAAATAAAA	1619
Qy	343	AACTTCGCAATCAAGCCTCAGGAGCGTCAATTTGTATCTGTCAACCTAGTGATGAGTCC	402
Db	1620	ATAAAGATGTAAAAGATTGGGAGCCAAATTACTCCACATGGGACACAGCTAGATGCA	1679
Qy	403	AGAGAAATGTGCAAAAGAGTTACAGAAGAACAGAGGCGATCATGTGTACATCCCAAC	462
Db	1680	GCATTAAAGAGGCACGTGTTTATTGCCAAGAAAAAAATTTGCTTTGTTCATCCTTTT	1739
Qy	463	CAGGAGCCTGCAGTGTAGCTGGACACAGGACAATTTGCCCTGGAAGTGTGTGAACCAAGTT	522
Db	1740	GACGACCTCTAACAAATGCTGGCCAAGGACATAGGACAAGAAATTCCTTGAAGATAAA	1799
Qy	523	CCTTTGGTGGAATGCACCTGGTGGTACCTGTAGTGGAGGAGGAATGCTTGGTGAATAGC	581
Db	1800	AATAAATTTTCATGTGCTCTTTGTTCCGGTGGGAGAGGAGGTATCTACGTGGTGATC	1858

```

RESULT 4
US-09-974-300-4749
; Sequence 4749, Application US/09974300
; Patent NO. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berkta, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4749
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Bacillus clausii
; US-09-974-300-4749

```

	Query Match	6.9%;	Score 70.6;	DB 10;	Length 954;
	Best Local Similarity	47.6%;	Prod. No. 6.8e-12;		
	Matches 248;	Conservative 0;	Mismatches 264;	Indels 9;	Gaps 1;
Qy	123	TCCTTTCTTCAAAATGTGAACCTCTCCAGAAAAACAGGATCTTTAAAGATTTCGTGGTGCCTCT	182		
Db	113	TGTTGGTTGAAATTAGAAACATTACAGCGCGGCTCATTTAAGCTCCGCGGGGCGAT	172		
Qy	183	CAATGCCGTGAGAGCTTGCTGTATGCTTTAGAAAGGAGCGGAAAAGCTGTGTGTAC	242		
Db	173	GAACGTGCTC-----TTGTCTATGTTCAGAAGAGTCAAGACAAATAAAGCGCTTGCCGC	223		
Qy	243	TCACAGCAGTGGAAAAACCATGCCAGGCTTCACCTATGCTGCCAAATTTGGAAGGAATTC	302		
Db	224	TTTTTCACGGAAACCATGCGCTTGGGTGCGCTACGCCGCCAAACAGAGCTCAAAACACA	283		
Qy	303	TGCTTATATTGTTGGTGCCTCCAGACAGCTCCAGACTGTAAAAAATTGCAATACAGCCTA	362		
Db	284	GGGCGACAAATTTTGTATCCGAAATCGGTACCGCTGCCAACTAGCGCGCTTTCAAGCGAG	343		
Qy	363	CGAGCGCTCAATTGTATACTGTGAACCTAGTGTATGATGCCAGAAAAATTTGCCAAAAG	422		
Db	344	CGGGGCAACAATAACCGTGACAGGGCGCAGCCCAAGATGAAGCAGGCGCACGCTGTATCCG	403		



## GENERAL INFORMATION:

APPLICANT: Gruys, Kenneth James  
APPLICANT: Mitsky, Timothy Albert  
APPLICANT: Kishore, Ganesh Murthy  
APPLICANT: Slater, Steven Charles  
APPLICANT: Padgett, Stephen Rogers  
APPLICANT: Stark, David Martin  
TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plasmids  
FILE REFERENCE: 11899.0155.DVUS02 (WOBT.155--3)  
CURRENT APPLICATION NUMBER: US/09/942,891  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: US 09/313,123  
PRIOR FILING DATE: 1999-05-17  
PRIOR APPLICATION NUMBER: US 08/673,388  
PRIOR FILING DATE: 1996-06-28  
PRIOR APPLICATION NUMBER: US 08/628,039  
PRIOR FILING DATE: 1996-04-04  
PRIOR APPLICATION NUMBER: US 08/614,877  
PRIOR FILING DATE: 1996-03-13  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent in version 3.1  
SEQ ID NO: 1  
LENGTH: 1545  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-942-891-1

Query Match 5.6%; Score 57; DB 9; Length 1545;

Best Local Similarity 43.6%; Pred. No. 2.3e-07;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

QY 216 AGAAGGAGCGGCGGAGGCTGTGTTACTCACAGCAGTGGAACCATGGCCAGGCTCTCAC 275  
DB 225 AGACAGAAAGCGCGGCTGATCAGCTCTCTCGGGTAAACAGCGCGCGGCTGCG 284  
QY 276 CTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATTGTGTGCGCCACAGACAGCTCCAGA 335  
DB 285 GTTTTCTTCTGCGCGGTAGGCGGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344  
QY 336 CTGTAAAAAAGCTTGAACACAGCTACGGAGCGTCAATTTGTTACTGTGAACCTAGTGA 395  
DB 345 CATCAAGTCGACCGGCTGCGCGGCTTCGGCGGCAAGTGTCTGCCAGCGCGCAACTT 404  
QY 396 TGAGTCCAGAGAAATTTGCAAAAAGAGTTACAGAAAGAAACAGAGGATCATGTGTACA 455  
DB 405 TGATGAGCGAAACCGCAAGCGATCGAAGTGTACAGAGAGGAGGTTTCACTGGGTGCC 464  
QY 456 TCCCAACAGGAGCCTGCGAGTGATAGCTGGACAAGGACAATTTGCCCTGGAAAGTGTGAA 515  
DB 465 GCGGTTGACCATCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 524  
QY 516 CCAGGTTCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 575  
DB 525 GCAGGAGCGCCATCTCGACCGCGTATTTGTGCGAGTGGCGGCGGCTGCTGGCTGCTGG 584  
QY 576 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTCTC 635  
DB 585 CTGCGGGTGTGATCAACAACACTGATGCCGCAATCAAGTGTGCGGTGAGAGCGGA 644  
QY 636 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTCTC 695  
DB 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATGCGGCTGATCGGTTGATCTGCGCGCGGT 704  
QY 696 TCCGAAACCATAGCAGATGGTGTCAATCCAGAGATTCGCTTGAACACTGCGCTTATATCC 755  
DB 705 AGGGTATTTGCTGAAGCGGTAGCGGTAAACGCGATCGGTGACGAAACCTTCCCGTTATG 764  
QY 756 CAGGAGCTTTGGTGGTATATCTTCACTGTCTACAGAGGATGAAT 800  
DB 765 CCAGAGATCTCGACGACATCATCACCGTCGATAGATGCGAT 809

## RESULT 8

US-09-942-891-5  
Sequence 5, Application US/09942891  
Publication No. US20030028917A1  
GENERAL INFORMATION:  
APPLICANT: Gruys, Kenneth James  
APPLICANT: Mitsky, Timothy Albert  
APPLICANT: Kishore, Ganesh Murthy  
APPLICANT: Slater, Steven Charles  
APPLICANT: Padgett, Stephen Rogers  
APPLICANT: Stark, David Martin  
TITLE OF INVENTION: Methods of Optimizing Substrate Pools and Biosynthesis of Poly-beta-hydroxybutyrate-co-poly-beta-hydroxyvalerate in Bacteria and Plasmids  
FILE REFERENCE: 11899.0155.DVUS02 (WOBT.155--3)  
CURRENT APPLICATION NUMBER: US/09/942,891  
CURRENT FILING DATE: 2002-02-01  
PRIOR APPLICATION NUMBER: US 09/313,123  
PRIOR FILING DATE: 1999-05-17  
PRIOR APPLICATION NUMBER: US 08/673,388  
PRIOR FILING DATE: 1996-06-28  
PRIOR APPLICATION NUMBER: US 08/628,039  
PRIOR FILING DATE: 1996-04-04  
PRIOR APPLICATION NUMBER: US 08/614,877  
PRIOR FILING DATE: 1996-03-13  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent in version 3.1  
SEQ ID NO: 5  
LENGTH: 1545  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-942-891-5

Query Match 5.6%; Score 57; DB 9; Length 1545;

Best Local Similarity 43.6%; Pred. No. 2.3e-07;

Matches 255; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

QY 216 AGAAGGAGCGGCGGAGGCTGTGTTACTCACAGCAGTGGAACCATGGCCAGGCTCTCAC 275  
DB 225 AGACAGAAAGCGCGGCTGATCAGCTCTCTCGGGTAAACAGCGCGCGGCTGCG 284  
QY 276 CTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATTGTGTGCGCCACAGACAGCTCCAGA 335  
DB 285 GTTTTCTTCTGCGCGGTAGGCGGTGAAGGCCCTGATCGTTATGCCAACGCCACCGCGGA 344  
QY 336 CTGTAAAAAAGCTTGAACACAGCTACGGAGCGTCAATTTGTTACTGTGAACCTAGTGA 395  
DB 345 CATCAAGTCGACCGGCTGCGCGGCTTCGGCGGCAAGTGTCTGCCAGCGCGCAACTT 404  
QY 396 TGAGTCCAGAGAAATTTGCAAAAAGAGTTACAGAAAGAAACAGAGGATCATGTGTACA 455  
DB 405 TGATGAGCGAAACCGCAAGCGATCGAAGTGTACAGAGAGGAGGTTTCACTGGGTGCC 464  
QY 456 TCCCAACAGGAGCCTGCGAGTGATAGCTGGACAAGGACAATTTGCCCTGGAAAGTGTGAA 515  
DB 465 GCGGTTGACCATCCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 524  
QY 516 CCAGGTTCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 575  
DB 525 GCAGGAGCGCCATCTCGACCGCGTATTTGTGCGAGTGGCGGCGGCTGCTGGCTGCTGG 584  
QY 576 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTCTC 635  
DB 585 CTGCGGGTGTGATCAACAACACTGATGCCGCAATCAAGTGTGCGGTGAGAGCGGA 644  
QY 636 AATGCAATTTACAGTTAAGGCTCTGAAACCTAGTGTGAAGGTATATGCTGCTGAACCTCTC 695  
DB 645 AGACTCCGCGCTGCTGAAAGCAGCGCTGGATGCGGCTGATCGGTTGATCTGCGCGCGGT 704  
QY 696 TCCGAAACCATAGCAGATGGTGTCAATCCAGAGATTCGCTTGAACACTGCGCTTATATCC 755  
DB 705 AGGGTATTTGCTGAAGCGGTAGCGGTAAACGCGATCGGTGACGAAACCTTCCCGTTATG 764







```
QY 475 GTGATAGCTGGACAAGGACAAATTCGCCGTGAAGTCTGTAACACAGGTTCTCTTTGGTGGAT 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410455 ATCATCGCGCGGACGACCGGTCGCTCTTGAGATCGTTGAGCGGATCGCGACGTCGCG 410514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 535 GCACCTGGTGGTACCTGAGTGAGGAGGAGGAATGCTTGGTGGAAATACAAATACAGTTAAG 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410515 ATGGTACTGCTTCCACTGCGGGTGGCGGCTTGGCTGCAGGCGTTGCAGCAGCGGTGAAG 410574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 595 GCTCTCAAAACCTAGTGTGAAGGTATATGCTG 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410575 GCACCTGCGGCTCATCGGAGGATCATCGGTG 410605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-974-300-4809
; Sequence 4809, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4809
; LENGTH: 926
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(926)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4809

Query Match 4.7%; Score 48.4; DB 10; Length 926;
Best Local Similarity 44.5%; Pred. No. 9.5e-05;
Matches 383; Conservative 0; Mismatches 456; Indels 21; Gaps 4;

QY 31 GATGTTGAAAAGCTCATATCAACATTCAGATTCATCCACTACACACGATGCTAAACA 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 GATATATTTCGAGCAAAATCAACAAATTAAGATGTGTCGACGCATACACCTTTGCAGAAA 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 91 AGCTCCATTTTGAATCAACTAACAGGCGCAATCTTTCTTCAAAATGTGAAGTCTTCCAG 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 GATCAGGTTTTGCTGAACGATATGATTCACAGTTTACTTAAACGGGAAGACTTGCAG 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 151 AAACAGGATCTTTTAAGATTGCTGTGCTCAATGCGGTGAGAGCTTGGTTCTGAT 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 GTTGTCAGATCTTTTAAATTCGCGCGC-----CTATTACCAAAATTTCTTCTGCTAT 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 211 GCTTTAGAAAGCAAGCAAGCTGTTGTACTCAGCAGTGGAAACCATGCCAGGCT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CAAAAGAAAGTTCGCGCTGTCG- GTTGTATGCGAAGCGCTGGAAACCATCACAAGG 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 271 CTCACCTATGCTCCCAAAATTTGAAGGAATTCCTGCTTATATTGTTGGTCCCGCAGACGT 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GTTGCTATTATCGCGCGCTTTAAAGTCAAGGTGTTATTTTATGCGGACTACCAG 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 331 CCAGACTGTAAAAAATTCGAATACAGCCTACGGAGCGTCAATTGTATATCTGTAACCT 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 CCAAGCAAAAAGTCGACAAAGTGAATTTTGGCAGGAGTATGCTAGACGTCAGGTTA 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 391 AGTGATGATCCAGAGAAATGTT-----GCAAAAGAGTTACAGAAACAGAA 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 ATTGGCGATACGTTTGTATGATTCTTATGCGGAGCGCAATTCATATTGCGAAGAACAGAA 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 442 GGCAATCATGTTACATCCCAACACAGGACCTGCTGAGTGTAGCTGACAGGACCAATTTGCC 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 ATGACGTTTCATCATTCATTTAACAGAGCAAAAGTGTATTCGCGGGAAGGAACAGTCGGA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 CTGGAAGTGTGA---ACCAGGTTCCCTTTGGTGGATGCACTGGTGGTGTACCTGTAGGTGGA 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 CTTGAAATTTATGAATGACATCGAGGAACGCCAGATTTTGTGTTTTCATCAATCGCGCGC 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 GGAGGAATGCTTCTGCTGGAATAGCAATTAAGTTAAGGCTCTGAAACCTAGTGTGAAGGTA 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GGTGGGCTTATTAGTGGGATGGCCACGTATATATAAAAGCGTTAGCCCCAACCTACGAAATG 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 TATGCTGCTGAACCCCTCAAAATGCAGATGACTGTACCAGTCCAAAGCTGGAAGGGAACCTG 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 ATTGGTTGTGAACCTCTGCTGCTGTCATCCATGACGAATCAATTAAGCAAGCAACGTT 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 679 ATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTGTCAAAATCCAGCATTTGGCTTG 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 GTTGAGTTAGACGAAATCGACAAATTTGTCGATGGAGCGGCTGTGAAAAAAGTCGCGGAC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 739 AACACCTGGCCTATTATCAGGGACCTTGTGGATGATATCTTCACTGTCACAGAGGATGAA 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 AAAACACTGGAATTTGCAAAAGAACTGCTAGATGACATTAATTGTTGTGCCAAGAGCCAAA 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 799 ATTAAGTGTCAACCCAGCTGTTGGGAGAGGATGAAACCTACTCTTGAACCTACAGCT 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ATTTGCACGACGATTTTAAACCTTTACATGAGAAATGCGCTAGTTGCCGACGG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 859 GGTGTTGGAGTGGCTGCTGT 878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 GCTATGCTTATTGCGGCTCT 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-781-986A-102
; Sequence 102, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```

; TOPOLOGY: linear
US-08-781-986A-102

Query Match          4.5%; Score 46.2; DB 7; Length 15249;
Best Local Similarity 45.2%; Pred. No. 0.0031;
Matches 364; Conservative 0; Mismatches 408; Indels 33; Gaps 4;

QY 31 GATGTTGAAAAGCTCATATCAACATTCGAGATTCATCCACCTCACCAGCTGCTAACCA 90
Db 13181 GATATCGATGAGCATTTTAAAGACTTTAAAGATATTGCAAGAAACACCTTTTACAATTA 13240

QY 91 AGCTCCATTTGAATCAACTAACAGGCGCAATCTTTCTTCAATGTGAACCTTCCAG 150
Db 13241 GACCATTAATCTCAAAAGTATGATTTAAAGCTTATTTAAACAGAGAGATTACAA 13300

QY 151 AAAACAGGATCTTTTAAGATTCGTGGTCTCTCAATGCCGTCAGAACCTTGGTTCCTGAT 210
Db 13301 TGGGTACGTCTTTTAATTAATTAAGAGGTGCTTACACGCTATTTCTGTTTATCAGATGAA 13360

QY 211 GCTTTAGAAAGGAGCGCAAGCTGTTGTTACTCACAGAGTGGAACACCATGGCCAGGCT 270
Db 13361 GCT-----AAAAGTAAAGGTATTACATGTGCAAGTCAGGTAATCATGCTCAAGGT 13411

QY 271 CTCACCTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATGTGGTGGCCCCAGACAGCT 330
Db 13412 GTTGCCTATACAGCTAAACAACTTAATTTAAACGCTGTTTATCTTTATGCGCAGTCACTACA 13471

QY 331 CCAGACTGTAAAAAACTTGCATTAACAGCCTACGGAGCGTCAATGTGTATCTGTGACCT 390
Db 13472 CCITTACAAAGGTAAATCAAGTAAAGTCTTTGGAAATAGTAACTGTTGAAGTTGTACT 13531

QY 391 AGTGATGAGTCCAGAGAAAA-----TGTTGCAAAAAGAGTTACAGAGAAGAAACAGAA 441
Db 13532 ACTGGTGATACATTTGATCACTGTTTAGCTGAAGCTTTAACTTATACAACTGAACATCAA 13591

QY 442 GGATCATGTATACATCCCAACAGGAGCCTGCGAGTGATAGCTGCAGCAAGGAGCAATGGCC 501
Db 13592 ATGAACCTTTATAGATCCATTTCAATAATGTTTCATACAAATTTCTGGACAAGGTAGCTTGC 13651

QY 502 CTGGAAGTGTGAACAGG-----TTCCCTTTGGTGGTGCACCTGGTGTACCT 549
Db 13652 AAAGAATGCTAGACAGCAAGCAAGCTGCAATGTTAACTTTGATTTATCTTTGCGGCA 13711

QY 550 GTAGGTGGAGGAGGAATGCTTGTGGAATAGCAATACAGTTAAGGCTCTGAAACCTAGT 609
Db 13712 ATTGGTGGTGGCGGTTAAATTTCAAGGTATTAGTACTTACTTTAAACCTATTCACTACC 13771

QY 610 GTCAAGGTATATCTGCTGAACCTCAATTCGAGATGACTGCTACAGCTCCAGCTGAAG 669
Db 13772 ACGAAATTTATAGGTGTTGAACCTTCAGGTGCAAGTAGTATGATGAATCTGTTGGTA 13831

QY 670 GGGAAACTGATGCCCAATCTTTATCTCCAGAAACCATAGCAGATGGTG---TCAAAATCC 726
Db 13832 AATAATCAGGTAGTACATTCGCTATATATCGATAAATTTGTCAGCGTGCATCTGAGCT 13891

QY 727 AGCATGGCTTGAACACCTGGCCCTATTATCAGGACCTTTGTGGATGATATCTTCACTGTC 786
Db 13892 AGATTTGGCGATATTACATTTGAAATTTGCAAAAGAAATGTCAGATGATTACGTTCAAGTA 13951

QY 787 ACAGAGGATCAAAATTAAGTGTGCAA 811
Db 13952 GATGAAGGTGCAGTTTGTGTTCTAGGA 13976

RESULT 15
US-09-939-980-207/c
; Sequence 207, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
;            Burdham, Martin
;            Hodgson, John
;            Knowles, David
;            Lonetto, Michael
;            Nicholas, Richard
;            Pratt, Julie
;            Reichard, Richard
;            Rosenberg, Martin
;            Ward, Judith

TITLE OF INVENTION: NO. US20020082234A1el Prokaryotic Polynucleotides,
Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,980
FILING DATE: 27-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/936,165
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 2698 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 207:
US-09-939-980-207

Query Match          4.5%; Score 46; DB 10; Length 2698;
Best Local Similarity 46.5%; Pred. No. 0.0011;
Matches 231; Conservative 0; Mismatches 248; Indels 18; Gaps 2;

QY 31 GATGTTGAAAAGCTCATATCAACATTCGAGATTCATCCACCTCACCAGTCTTAAACA 90
Db 575 GATATCGATGAAGACATTTTAAAGACTTAAAGATATTGTCAAAGAAACACCTTTACAATTA 516

QY 91 AGCTCCATTTGTAATCAACTAACAGGCGCAATCTTTCTTCAAAATGTGAACCTCTTCCAG 150
Db 515 GACCATTAATCTCAAAAGTATGATGTTAAAGTTTATTTAAACAGAGAGATTACAA 456

QY 151 AAAACAGGATCTTTTAAAGATTCGTGGTCTCTCAATGCGCTCAGAGCTGTTGCTGAT 210
Db 455 TGGGTACGTTCTTTTAAATTAAGAGGTGCTTACACGCTATTTCTGTTTATCAGATGAA 396

QY 211 GCTTTAGAAAGGAGCGCAAGCTGTTGTTACTCACAGAGTGAACCATGCCAGGCT 270
Db 395 GCT-----AAAAGTAAAGGTATTACATGTCGAGGTGCGAGGTATCATGCTCAAGGT 345

QY 271 CTCACCTATGCTGCCAAATTTGGAAGGAATTCCTGCTTATATTTGTTGGTGGCCCCAGACAGCT 330
Db 344 GTTGCCTATACAGCTAAACAACTTAATNTAAAGCTGTTATCTTTATGCCAGTCACTACA 285

QY 331 CCAGACTGTAAAAAACTTGCATTAACAGCCTACGGA-----CGGTCAATTGTATAC 381
Db 284 CCTTNACAAAGGTAATAAAGTAAAGTTCTTTTGGAAATAGTAACGTTGAAGTTGTACTC 225
```

QY	382	TGTGAACCTAGTGATCGAGTCCAGAGAAAATGTTGCAAAAAGAGTTACAGAAGAAACAGAA	441
Db	224	ACTGGTGATACATTTCGATCNCCTTTAGCTGAAGCTTTAACTTTATACAAGTGAACATCAA	165
QY	442	GGCATCATGGTACATCCCAACCAAGGAGCCTGCAGTGATAGCTGGACAAGGGACNATTGCC	501
Db	164	ATGAACCTTTATAGATCCCATTCATAATGTTTCATACAAATTTCTGGACAAGGTACGCTTGCT	105
QY	502	CTGGAAGTGCTGAACCA	518
Db	104	AAAGAAATGCTAGACA	88

Search completed: June 24, 2003, 09:40:02  
Job time : 1742 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:35:44 ; Search time 46.0677 Seconds  
(without alignments)  
983.448 Million cell updates/sec

Title: US-09-889-609B-10

Perfect score: 1735

Sequence: 1 MCAQVCISFADVEKAHINR.....SSITWVKQAEKRPASTQSVSV 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1735	100.0	340	21	AA07734
2	1735	100.0	340	22	AAU09124
3	1735	100.0	340	22	AAE08342
4	1735	100.0	340	22	AAE94477
5	1731	99.8	340	22	AAU50262
6	1582.5	99.5	340	22	AAE78808
7	1582.5	91.2	339	21	AA07731
8	1062.5	61.2	228	23	ABB89713
9	450	25.9	469	22	ABB63708
10	414.5	23.9	1181	22	ABG24298

11	368	21.2	502	20	AA32941	Mutant threonine d
12	368	21.2	502	20	AAV05705	Feedback insensiti
13	368	21.2	532	20	AAV32943	Mutant threonine d
14	368	21.2	532	20	AAV05707	Feedback insensiti
15	368	21.2	539	20	AAV32942	Mutant threonine d
16	368	21.2	539	20	AAV05706	Feedback insensiti
17	368	21.2	545	20	AAV32947	Mutant threonine d
18	368	21.2	545	20	AAV05711	Feedback insensiti
19	368	21.2	592	20	AAV32939	Mutant threonine d
20	368	21.2	592	20	AAV32951	Wild type threonin
21	368	21.2	592	20	AAV05702	Arabidopsis wild-t
22	368	21.2	592	20	AAV05703	Feedback insensiti
23	368	21.2	600	20	AAV32952	Mutant threonine d
24	368	21.2	609	20	AAV32940	Mutant threonine d
25	368	21.2	609	20	AAV05704	Feedback insensiti
26	367	21.2	424	22	AA81976	S. epidermidis ope
27	367	21.2	424	23	ABP39031	Staphylococcus epi
28	363	20.9	590	20	AAV32950	Mutant threonine d
29	363	20.9	592	20	AAV32948	Mutant threonine d
30	361.5	20.8	422	23	ABB48174	Listeria monocytog
31	351.5	20.3	416	23	ABB54559	Lactococcus lactis
32	349.5	20.1	310	22	AA90837	C glutamicum prote
33	349.5	20.1	310	22	AA879703	Corynebacterium gl
34	348.5	20.1	441	15	AA54223	L.lactis branched
35	312.5	18.0	436	16	AA94690	Threonine dehydrat
36	311.5	18.0	436	22	AA92074	C glutamicum prote
37	311	17.9	621	22	ABG30299	Novel human diagno
38	310.5	17.9	436	16	AA94686	Threonine dehydrat
39	309.5	17.8	423	22	AA879762	Corynebacterium gl
40	309.5	17.8	436	16	AA94687	Threonine dehydrat
41	309.5	17.8	436	16	AA94688	Threonine dehydrat
42	307.5	17.7	436	16	AA94689	Threonine dehydrat
43	307.5	17.7	436	16	AA94691	Threonine dehydrat
44	305	17.6	316	22	ABB66014	Drosophila melanog
45	279.5	16.1	340	22	AAU23238	Novel human enzyme

ALIGNMENTS

RESULT 1	
AA07734	
ID	AA07734 standard; Protein; 340 AA.
XX	
AC	AA07734;
XX	
DT	07-NOV-2000 (first entry)
XX	
DE	Amino acid sequence of a human serine racemase polypeptide.
XX	
KW	Serine racemase; N-methyl-D-aspartate receptor; neural death;
KW	neural dysfunction; NMDA receptor; Parkinson's disease;
KW	Huntington's disease; motor neurone disease; Alzheimer's disease.
XX	
OS	Homo sapiens.
XX	
PN	WO200043526-A1.
XX	
PD	27-JUL-2000.
XX	
PF	18-JAN-2000; 2000WO-US00938.
XX	
PR	19-JAN-1999; 99US-0116333.
PR	21-JUL-1999; 99US-0144839.
XX	
XX	28-JUL-1999; 99US-0145953.
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX	
PI	Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;
PI	Ferris CD;
XX	
XX	WPI; 2000-482915/42.
DR	N-PSDB; AAA59299.



XX AAE08342;  
 AC 15-NOV-2001 (first entry)  
 XX Human pyridoxal-phosphate dependent enzyme 22406 protein.  
 DE  
 XX Human; pyridoxal phosphate dependent enzyme; nootropic; neuroprotective;  
 KW anticonvulsant; cerbroprotective; cardiant; vasotropic; gene therapy;  
 KW epilepsy; Alzheimer's disease; chromosome 17; serine racemase; stroke;  
 KW behavioural change; neurodegenerative disorder; schizophrenia; atresia;  
 KW rheumatic heart failure; circulatory disorder; hepatic injury; jaundice;  
 KW lung disorder; nodular hyperplasia; stenosis; skeletal muscle disorder;  
 KW tumour; rhabdomyosarcoma; dermal fibroblast disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Location/Qualifiers  
 FT Modified-site 8..11  
 FT Domain /label= Casein\_kinase\_II\_phosphorylation\_site  
 FT 19..315  
 FT /note= "Pyridoxal-phosphate dependent enzyme family  
 FT domain"  
 FT Modified-site 38..40  
 FT /label= Protein\_kinase\_C\_phosphorylation\_site  
 FT Modified-site 47..60  
 FT /note= "Serine/threonine dehydratases pyridoxal-phosphate  
 FT attachment site"  
 FT Modified-site 54..56  
 FT /label= Protein\_kinase\_C\_phosphorylation\_site  
 FT Modified-site 59..64  
 FT /label= N\_myristoylation\_site  
 FT Modified-site 88..93  
 FT /label= N\_myristoylation\_site  
 FT Modified-site 109..112  
 FT /label= Casein\_kinase\_II\_phosphorylation\_site  
 FT Modified-site 140..143  
 FT /note= "cAMP and cGMP-dependent protein kinase  
 FT phosphorylation site"  
 FT Domain 176..197  
 FT /label= Transmembrane\_domain  
 FT Modified-site 187..192  
 FT /label= N\_myristoylation\_site  
 FT Modified-site 196..198  
 FT /label= Protein\_kinase\_C\_phosphorylation\_site  
 FT Modified-site 203..205  
 FT /label= Protein\_kinase\_C\_phosphorylation\_site  
 FT Modified-site 212..215  
 FT /label= Casein\_kinase\_II\_phosphorylation\_site  
 FT Modified-site 235..238  
 FT /label= Casein\_kinase\_II\_phosphorylation\_site  
 FT Modified-site 239..244  
 FT /label= N\_myristoylation\_site  
 FT Modified-site 261..264  
 FT /label= Casein\_kinase\_II\_phosphorylation\_site  
 FT Modified-site 287..292  
 FT /label= N\_myristoylation\_site  
 FT Domain 308..326  
 FT /label= Transmembrane\_domain  
 XX  
 PN WO200160987-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 XX 20-FEB-2001; 2001WO-US05365.  
 PF  
 XX 17-FEB-2000; 2000US-0183208.  
 PR  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX Meyers RA, Rudolph-Owen LA;  
 PI  
 XX WPI; 2001-529909/58.  
 DR

DR N-PSDB: AAD14461.  
 XX Novel polypeptide of the human pyridoxal phosphate dependent family  
 PT useful in screening and detection assays and for treatment, e.g. of  
 PT epilepsy and Alzheimer's  
 XX Claim 8; Fig 1; 12lpp: English.  
 PS  
 XX The present sequence is human pyridoxal phosphate dependent enzyme  
 CC 22406 which is a serine racemase. Human 22406 gene is located on  
 CC chromosome 17 between D17S849 and D17S796. The protein 22406  
 CC is a modulator of D-serine. D serine has been shown to modify  
 CC behavioural changes associated with learning, memory and convulsions.  
 CC Human 22046 and compounds that modulate the expression or activity are  
 CC used to treat or diagnose neurodegenerative disorders including  
 CC Alzheimer's disease, schizophrenia as well as quell anxiety and  
 CC epilepsy and prevent damage from stroke as well as cardiac (heart  
 CC failure, rheumatic heart failure) and circulatory disorders, liver  
 CC disorders (hepatic injury, jaundice), lung disorders, prostate  
 CC disorders (benign enlargement, nodular hyperplasia), colon disorders  
 CC (atresia, stenosis), skeletal muscle disorders (tumours-rhabdomyosarcoma)  
 CC and dermal fibroblast disorders. Human 22406 cDNA is also useful in gene  
 CC therapy.  
 XX  
 SQ Sequence 340 AA;  
 Query.Match 100.0%; Score 1735; DB 22; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-167;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCAQYCIISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKGTGSKIRGA 60  
 DB 1 MCAQYCIISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKGTGSKIRGA 60  
 QY 61 LNAVRSYLPDALERPKAVVTHSSNGHGOALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 DB 61 LNAVRSYLPDALERPKAVVTHSSNGHGOALTYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNPEPAVIAQGGTIALEVLNQVPLVDAL 180  
 DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNPEPAVIAQGGTIALEVLNQVPLVDAL 180  
 QY 181 VVPVGGGMLAGIAITVRALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 DB 181 VVPVGGGMLAGIAITVRALKPSVKVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 QY 241 KSSIGLNTWPIIRDLVDVDFITVTEDEIKCATOLVHERMKLLIETPTAGVGVAAVLSQHFT 300  
 DB 241 KSSIGLNTWPIIRDLVDVDFITVTEDEIKCATOLVHERMKLLIETPTAGVGVAAVLSQHFT 300  
 QY 301 VSPEVKNICIVLSGGNVDLTSSITVWQKERPASYSQSVSV 340  
 DB 301 VSPEVKNICIVLSGGNVDLTSSITVWQKERPASYSQSVSV 340  
 RESULT 4  
 AAB94477  
 ID AAB94477 standard; Protein; 340 AA.  
 XX  
 AC AAB94477;  
 XX  
 XX 26-JUN-2001 (first entry)  
 DT  
 XX Human protein sequence SEQ ID NO:15149.  
 DE  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW  
 XX Homo sapiens.  
 OS  
 XX EP1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX

PF 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 XX WPI: 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 15149; 2537pp + CD ROM; English.  
 PS  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX Sequence 340 AA;  
 SQ  
 Query Match 100.0%; Score 1735; DB 22; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-167;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MCAQYICISFADVEKAHINIRDSIHITPVLTSILNQLTGRNLFKCELFQKTSKIRGA 60  
 Db 1 MCAQYICISFADVEKAHINIRDSIHITPVLTSILNQLTGRNLFKCELFQKTSKIRGA 60  
 QY 61 LNAVRLVPDALERPKAVVTHSSNGHQAITYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 Db 61 LNAVRLVPDALERPKAVVTHSSNGHQAITYAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
 QY 121 YGASTVYCEPDSERENAKRVTEETEGIMVHPNQPENAVIAGQGTIALEVLNQVPLDAL 180  
 Db 121 YGASTVYCEPDSERENAKRVTEETEGIMVHPNQPENAVIAGQGTIALEVLNQVPLDAL 180  
 QY 181 VVPVGGGMLAGIATTVKALPSVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 Db 181 VVPVGGGMLAGIATTVKALPSVYAAEPSNADDCYOSKLGKMLPNLYPPETIADGV 240  
 QY 241 KSSIGLNTWPIIRDLVDVDFTEDEIKCATOLVWERMKLLIEPTAGVGVAAVLVLSQHFT 300  
 Db 241 KSSIGLNTWPIIRDLVDVDFTEDEIKCATOLVWERMKLLIEPTAGVGVAAVLVLSQHFT 300  
 QY 301 VSPKVNICIVLGGNVDLTSSITVWKAERPASTQSVSV 340  
 Db 301 VSPKVNICIVLGGNVDLTSSITVWKAERPASTQSVSV 340

Db 301 VSPKVNICIVLGGNVDLTSSITVWKAERPASTQSVSV 340  
 RESULT 5  
 AAM50262  
 ID AAM50262 standard; Protein; 340 AA.  
 AC AAM50262;  
 XX 21-JAN-2002 (first entry)  
 DT Human serine racemase.  
 DE  
 XX Serine racemase; human; D-serine; regulation;  
 KW glutamate N-methyl-D-aspartate receptor; neurodegenerative disease;  
 KW stroke; Alzheimer's disease; Parkinson's disease; pain; myopathy;  
 KW nootropic; neuroprotective; cerebroprotective; antiparkinsonian;  
 KW analgesic; diagnosis; therapy; screening.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Modified-site 41..59  
 FT /note= "prosite serine/threonine dehydratase  
 FT pyridoxal-phosphate attachment site"  
 FT Modified-site 71..128  
 FT /note= "serine/threonine dehydratase  
 FT pyridoxal-phosphate attachment site"  
 FT Modified-site 154..169  
 FT /note= "cysteine synthase/cystathione beta-synthase  
 FT attachment site"  
 FT  
 PN WO200173077-A2.  
 PD 04-OCT-2001.  
 XX 30-MAR-2001; 2001WO-EP03668.  
 XX 31-MAR-2000; 2000US-193748P.  
 PR 03-APR-2000; 2000US-194249P.  
 PA (FARB ) BAYER AG.  
 XX Ramakrishnan S;  
 XX WPI: 2001-648444/74.  
 DR N-PSDB; AAI70575.  
 XX Polynucleotide encoding serine racemase enzyme and the enzyme useful  
 PT for screening reagents regulating the activity of the enzyme in a  
 PT neuron disease caused by over- or under-activation of glutamate  
 PT N-methyl-D-aspartate  
 XX Claim 1; Fig 2; 66pp; English.  
 XX The present sequence is that of human serine racemase, an enzyme  
 CC that catalyses the conversion of L-serine to D-serine. Neuron  
 CC damage following various nervous system diseases is often caused by  
 CC activation of glutamate N-methyl-D-aspartate (NMDA) receptors in  
 CC the brain. This receptor is activated by the binding of D-serine.  
 CC Regulation of D-serine levels through regulation of serine racemase  
 CC may therefore prevent or minimise neuron damage in neurogenic and  
 CC myopathic disorders, neurodegenerative disorders such as Alzheimer's  
 CC disease and Parkinson's disease, and disorders leading to peripheral  
 CC and chronic pain. The invention provides serine racemase  
 CC polypeptides and methods for their recombinant production. These  
 CC polypeptides are used in claimed methods of screening for agents  
 CC that modulate or decrease serine racemase activity. A claimed  
 CC pharmaceutical composition comprising either an expression vector  
 CC that contains a serine racemase polynucleotide, or a reagent that  
 CC modulates serine racemase enzyme activity, is used to modulate  
 CC serine racemase activity in a disease, particularly neuron damage or  
 CC a neurodegenerative disease caused by the over- or under-activation



CC of the glutamate NMDA receptor.

XX Sequence 340 AA;  
SQ  
Query Match 99.8%; Score 1731; DB 22; Length 340;  
Best Local Similarity 99.7%; Pred. No. 4.1e-167;  
Matches 339; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFKTSFKIRGA 60  
DB 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFKTSFKIRGA 60  
QY 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTVAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
DB 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTVAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAAGQGTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAAGQGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLDVDDIETVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
DB 241 KSSIGLNTWPIIRDLDVDDIETVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
QY 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340  
DB 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340

## RESULT 6

AAG78808  
ID AAG78808 standard; Protein; 340 AA.

XX AC AAG78808;

XX DT 27-NOV-2001 (first entry)

XX DE Serine/threonine dehydrase 37.

XX KW Serine/threonine dehydrase 37; cancer; HIV infection; anti-HIV;  
XX cytosolic.

XX OS Unidentified.

XX PN CNL300824-A.

XX PD 27-JUN-2001.

XX PF 21-DEC-1999; 99CN-0125662.

XX PR 21-DEC-1999; 99CN-0125662.

XX PA (YFU-) UNIV FUDAN.

XX PI Mao Y, Xie Y;

XX DR WPI: 2001-530468/59.

XX DR N-PSDB: AAI65020.

XX PT Polypeptide-serine/threonine dehydrase 37 and polynucleotide for coding  
XX this polypeptide -

XX PS Claim 1; Page 27 (Disclosure); 34pp; Chinese.

XX CC The present sequence is the protein sequence for serine/threonine  
XX dehydrase 37. The dehydrase and its coding sequence are useful for  
XX treating diseases e.g. cancer and HIV infection.

XX SQ Sequence 340 AA;

Query Match 99.5%; Score 1727; DB 22; Length 340;  
Best Local Similarity 99.7%; Pred. No. 1e-166;  
Matches 339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFKTSFKIRGA 60  
DB 1 MCAQYCTSPADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKCELFKTSFKIRGA 60  
QY 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTVAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
DB 61 LNAVRSILVPDALERKPKAVVTHSSGNHGQALTVAAKLEGIPAYIVVPTAPDCKKLAIOA 120  
QY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAAGQGTIALEVLNQVPLVDAL 180  
DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNQPAPVIAAGQGTIALEVLNQVPLVDAL 180  
QY 181 VVPVGGGMLAGIAITVKALKPSVKVYAAPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
DB 181 VVPVGGGMLAGIAITVKALKPSVKVYAAPSNADDCYQSKLKGKLPNLYPPETIADGV 240  
QY 241 KSSIGLNTWPIIRDLDVDDIETVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
DB 241 KSSIGLNTWPIIRDLDVDDIETVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQT 300  
QY 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340  
DB 301 VSPEVKNICIVLGGNVDLTSSITWVKQAEPPASYSQSVS 340

## RESULT 7

AAB07731  
ID AAB07731 standard; Protein; 339 AA.

XX AC AAB07731;

XX DT 07-NOV-2000 (first entry)

XX DE Amino acid sequence of a murine serine racemase polypeptide.

XX KW Serine racemase; N-methyl-D-aspartate receptor; neural death;

XX KW neural dysfunction; NMDA receptor; Parkinson's disease;

XX KW Huntington's disease; motor neurone disease; Alzheimer's disease.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 337 /note= "Ser encoded by T in AAA59294"

XX FT WO200043526-A1.

XX PN 27-JUL-2000.

XX PD 18-JAN-2000; 2000WO-US00938.

XX PF 19-JAN-1999; 99US-0116333.

XX PR 21-JUL-1999; 99US-0144839.

XX PR 28-JUL-1999; 99US-0145953.

XX PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX PI Snyder SH, Wolosker H, Sheth K, Masaaki T, Mothet J, Brady RO;

XX PI Ferris CD;

XX DR WPI: 2000-482915/42.

XX DR N-PSDB: AAA59294, AAA59300.

XX PT Mammalian serine racemase preparations, used to identify modulators  
XX which can be used to treat diseases associated with  
XX N-methyl-D-aspartate receptor overactivation, e.g. Alzheimer's disease

XX PT

PS Claim 7; Page 48-49; 54pp; English.

XX The present sequence represents a mammalian serine racemase, which has  
 CC a specific activity of at least 0.003 micromole L-serine/mg/hour.  
 CC The enzyme catalyses the direct racemisation of L-serine to D-serine.  
 CC D-serine appears to be an endogenous ligand of N-methyl-D-aspartate  
 CC (NMDA) receptors. The mammalian serine racemases can be used to identify  
 CC modulators, which can be used in the treatment of acute or chronic  
 CC neural death or dysfunction mediated by overactivation of N-methyl-D-  
 CC aspartate (NMDA) receptors. Overactivation of the receptors is  
 CC associated with Parkinson's disease, Huntington's disease, motor neurone  
 CC disease and Alzheimer's disease.

XX Sequence 339 AA:

XX Query Match 91.2%; Score 1582.5; DB 21; Length 339;  
 CC Best Local Similarity 89.7%; Pred. No. 5.1e-152;  
 CC Matches 305; Conservative 19; Mismatches 15; Indels 1; Gaps 1;

OY 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 DB 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60

OY 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120  
 DB 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120

OY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180

OY 181 VVPVGGGMLAGIAITVKALPKSVKYVAEPNSNADDCYQSKLKGKLPNLYPPTIADGV 240  
 DB 181 VVPVGGGMLAGIAITVKALPKSVKYVAEPNSNADDCYQSKLKGKLPNLYPPTIADGV 240

OY 241 KSSIGLNTWPIIRDLDVDDIFTVTEDEIKCATQLWGRMKLLIEPTAGVGVAAVLSQHFQT 300  
 DB 241 KSSIGLNTWPIIRDLDVDDIFTVTEDEIKCATQLWGRMKLLIEPTAGVGVAAVLSQHFQT 300

OY 301 VSPKVKNICIVLGGNVDLTSSITWVKQAEKRPASYQSVV 340  
 DB 301 VSPKVKNICIVLGGNVDLTSSITWVKQAEKRPASYQSVV 340

RESULT 8  
 ABB89713  
 ID ABB89713 standard; Protein: 228 AA.

XX AC ABB89713;  
 XX DT 24-MAY-2002 (first entry)  
 XX DE Human polypeptide SEQ ID NO 2089.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.  
 XX WO2001190304-A2.  
 XX 29-NOV-2001.  
 XX 18-MAY-2001; 2001WO-US16450.  
 XX 19-MAY-2000; 2000US-205515P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;

XX WPI: 2002-122018/16.  
 DR N-PSDB; ABL90122.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders.

XX Claim 11; SEQ ID NO 2089; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL9449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

XX Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [fp.wipo.int/pub/published\\_pct\\_sequences](http://fp.wipo.int/pub/published_pct_sequences).

XX Sequence 228 AA:

XX Query Match 61.2%; Score 1062.5; DB 23; Length 228;  
 CC Best Local Similarity 93.4%; Pred. No. 2e-99;  
 CC Matches 213; Conservative 2; Mismatches 12; Indels 1; Gaps 1;

OY 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60  
 DB 1 MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGA 60

OY 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120  
 DB 61 LNAVSLVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIOA 120

OY 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180  
 DB 121 YGASIVYCEPSDESRENVAKRVTETEGIMVHPNOEPAVIAGOGTIALEVLNOVPLVDAL 180

OY 181 VVPVGGGMLAGIAITVKALPKSVKYVAEPNSNADDCYQSKLKGKLP 228  
 DB 181 VVPVGGGMLAGIAITVKALPKSVKYVAXXP-QMQMTVPVQAEGLMP 227

RESULT 9  
 ABB63708  
 ID ABB63708 standard; Protein: 469 AA.

XX AC ABB63708;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 17916.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.





CC R554H amino acid substitutions (numbering according to wild-type  
 CC TD) in the R4 and R6 regulatory regions. These forms of TD are not  
 CC only insensitive to feedback inhibition by isoleucine, but are also  
 CC insensitive to structural analogues of isoleucine that are toxic to  
 CC plants and microorganisms which synthesize only wild-type TD.  
 CC Nucleotide sequences encoding mutated forms of TD can therefore be  
 CC used to create cells that are insensitive to compounds normally  
 CC toxic to cells expressing only wild-type TD enzymes, and thus may  
 CC be used to provide a biochemical selectable marker. Transformsants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX  
 SQ Sequence 502 AA;  
 Query Match 21.2%; Score 368; DB 20; Length 502;  
 Best Local Similarity 30.2%; Pred. No. 2.8e-28;  
 Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;  
 QY 26 TPVLTSILNQLTGRLFFKCFELQGTGSKIRGALNAVRLVDPALERKPKAVVTHSSG 85  
 Db 21 SPLQLAKLSKRLGVRMYLKRDLQPVFSFKURGAYNMVKLPADQL---AKGVICSSAG 77  
 QY 86 NHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIQAYGASIVYCEPSDESRENVAKRVTTE 145  
 Db 78 NHAQGVALSASKLGCTAVIVMPVTTPETIKQAVENLGATVVLFGDSYDQAAHAK-IRAE 136  
 QY 146 TEGI-MVHPNQBPVAVIAGOGTIALEVLNQV--PLVDALVVPVGGGMLAGIAITVKALKP 202  
 Db 137 EGGLETFIPFDHPDVIAGOGTIVGMEITRQAKGPL-HAIFVPVGGGGLIAGIAIYKRVSP 195  
 QY 203 SVKYVAEFSNADDCYQSKLGLKMPNLYPPETI-----ADGVK-SSIGLNTWPIIR 253  
 Db 196 EVKIIGVEPADAN-----AMALSLHHGERVILDQVGGFADGVAVKEVEEETFRISR 246  
 QY 254 DLVDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLQSQHFQTVSPVKNICIVLS 313  
 Db 247 NLMGCVLWLTDRDAICASIKDMFEKRNITLPEAGALALAGA-EAYCKYYGLKDVNVVAITS 305  
 QY 314 GGNVDL-----TSSITVWVKQAPRPASYQ 336  
 Db 306 GANNMFDKLRIVTELANYGROQEAVALTMPEKPGSFK 343  
 RESULT 13  
 AAY32943  
 ID AAY32943 standard; Protein; 532 AA.  
 XX  
 AC AAY32943;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 XX  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN WO9941395-A1.  
 XX  
 PD 19-AUG-1999.  
 XX  
 PF 08-JAN-1999; 99WO-US00560.  
 XX  
 PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.

XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 PI  
 XX WPI: 1999-527375/44.  
 XX N-PSDB; AAZ11201.  
 XX  
 PT New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX  
 PS Claim 13; Page 111-114; 194pp; English.  
 XX  
 CC This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX  
 SQ Sequence 532 AA;  
 Query Match 21.2%; Score 368; DB 20; Length 532;  
 Best Local Similarity 30.2%; Pred. No. 3.1e-28;  
 Matches 102; Conservative 65; Mismatches 129; Indels 42; Gaps 10;  
 QY 26 TPVLTSILNQLTGRLFFKCFELQGTGSKIRGALNAVRLVDPALERKPKAVVTHSSG 85  
 Db 51 SPLQLAKLSKRLGVRMYLKRDLQPVFSFKURGAYNMVKLPADQL---AKGVICSSAG 107  
 QY 86 NHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIQAYGASIVYCEPSDESRENVAKRVTTE 145  
 Db 108 NHAQGVALSASKLGCTAVIVMPVTTPETIKQAVENLGATVVLFGDSYDQAAHAK-IRAE 166  
 QY 146 TEGI-MVHPNQBPVAVIAGOGTIALEVLNQV--PLVDALVVPVGGGMLAGIAITVKALKP 202  
 Db 167 EGGLETFIPFDHPDVIAGOGTIVGMEITRQAKGPL-HAIFVPVGGGGLIAGIAIYKRVSP 225  
 QY 203 SVKYVAEFSNADDCYQSKLGLKMPNLYPPETI-----ADGVK-SSIGLNTWPIIR 253  
 Db 226 EVKIIGVEPADAN-----AMALSLHHGERVILDQVGGFADGVAVKEVEEETFRISR 276  
 QY 254 DLVDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGVAAVLQSQHFQTVSPVKNICIVLS 313  
 Db 277 NLMGCVLWLTDRDAICASIKDMFEKRNITLPEAGALALAGA-EAYCKYYGLKDVNVVAITS 335  
 QY 314 GGNVDL-----TSSITVWVKQAPRPASYQ 336  
 Db 336 GANNMFDKLRIVTELANYGROQEAVALTMPEKPGSFK 373  
 RESULT 14  
 AAY05707  
 ID AAY05707 standard; Protein; 532 AA.  
 XX  
 AC AAY05707;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE Feedback insensitive mature threonine dehydratase/deaminase.  
 XX  
 KW Threonine dehydratase/deaminase; omr1 gene; feedback inhibition;  
 KW transgenic plant; selectable marker; isoleucine; mutant.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 04:18:31 ; Search time 21.5317 Seconds  
(without alignments)  
464.608 Million cell updates/sec

Title: us-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYCISFADVEKAHINIR.....SSITWVKAERPAASYQSVSV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1735	100.0	340	4	US-09-789-300A-2
2	367	21.2	424	4	US-09-134-001C-3876
3	348.5	20.1	441	1	US-08-403-866-10
4	310.5	17.9	436	3	US-08-669-378-2
5	310.5	17.9	436	3	US-08-669-378-12
6	309.5	17.8	436	3	US-08-669-378-4
7	309.5	17.8	436	3	US-08-669-378-6
8	309.5	17.8	436	3	US-08-669-378-10
9	307.5	17.7	436	3	US-08-669-378-8
10	283.5	16.3	378	4	US-09-789-300A-4
11	259.5	15.0	325	4	US-09-088-435-1
12	241.5	13.9	367	4	US-09-134-001C-4168
13	187	10.8	551	1	US-08-120-960-2
14	187	10.8	551	4	US-09-347-878-9
15	174	10.0	312	4	US-09-134-001C-3920
16	164.5	9.5	319	4	US-09-134-001C-3330
17	124.5	7.2	499	4	US-09-594-193-4
18	121.5	7.0	404	4	US-08-887-534A-34
19	120	6.9	497	4	US-09-594-193-2
20	106	6.1	487	4	US-09-594-193-13
21	105	6.1	540	4	US-08-973-462-22
22	105	6.1	1786	4	US-08-973-462-8
23	101.5	5.9	212	4	US-08-973-462-25
24	101.5	5.9	630	4	US-08-973-462-9
25	99.5	5.7	484	4	US-09-066-046-8
26	99	5.7	1642	2	US-08-662-227-2
27	99	5.7	1642	4	US-09-017-947-2

28	98.5	5.7	434	4	US-09-594-193-7	Sequence 7, Appli
29	98	5.6	403	4	US-09-134-001C-5236	Sequence 5236, Ap
30	93	5.4	1648	2	US-08-662-227-35	Sequence 35, Appl
31	93	5.4	1648	4	US-09-017-947-35	Sequence 35, Appl
32	92.5	5.3	619	4	US-09-066-046-2	Sequence 2, Appli
33	91	5.2	1582	2	US-08-404-531B-9	Sequence 9, Appli
34	91	5.2	1582	3	US-08-476-900A-9	Sequence 9, Appli
35	91	5.2	1582	3	US-08-488-546A-9	Sequence 9, Appli
36	91	5.2	1582	4	US-08-726-320-5	Sequence 5, Appli
37	91	5.2	1582	4	US-09-208-716-5	Sequence 5, Appli
38	89	5.1	1642	1	US-08-447-411-45	Sequence 45, Appl
39	88.5	5.1	388	4	US-09-134-001C-4951	Sequence 4951, Ap
40	88	5.1	395	2	US-08-404-531B-3	Sequence 3, Appli
41	88	5.1	395	3	US-08-476-900A-3	Sequence 3, Appli
42	88	5.1	395	3	US-08-488-546A-3	Sequence 3, Appli
43	88	5.1	1580	4	US-08-726-320-1	Sequence 1, Appli
44	88	5.1	1580	4	US-09-208-716-1	Sequence 1, Appli
45	88	5.1	1581	2	US-08-404-531B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-789-300A-2  
: Sequence 2, Application US/09789300A  
: Patent No. 6458576  
: GENERAL INFORMATION:  
: APPLICANT: Meyers, Rachel  
: APPLICANT: Rudolph-Owen, Laura A.  
: TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
: FILE REFERENCE: 35800/208926  
: CURRENT APPLICATION NUMBER: US/09/789.300A  
: CURRENT FILING DATE: 2001-02-20  
: PRIOR APPLICATION NUMBER: US 60/183,208  
: PRIOR FILING DATE: 2000-02-17  
: NUMBER OF SEQ ID NOS: 4  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 2  
: LENGTH: 340  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
: US-09-789-300A-2

Query Match	100.0%	Score 1735;	DB 4;	Length 340;
Best Local Similarity	100.0%	Pred. No. 2.1e-189;	Mismatches 0;	Indels 0;
Matches 340;	Conservative 0;			Gaps 0;
Qy	1	MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFORTGSFKIRGA	60	
Db	1	MCAQYCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFORTGSFKIRGA	60	
Qy	61	LNAVRSVPDALEKRPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA	120	
Db	61	LNAVRSVPDALEKRPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPTAPDCKKLAIOA	120	
Qy	121	YGASIVVCEPDESRENVAKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL	180	
Db	121	YGASIVVCEPDESRENVAKRVTEETEGIMVHPNQEPAVIAGOGTIALEVLNQVPLVDAL	180	
Qy	181	VVPVGGGMLAGIAITVKALKPSVKVYAAEPNADDCYQSKLKGKLPNLYPPETIADGV	240	
Db	181	VVPVGGGMLAGIAITVKALKPSVKVYAAEPNADDCYQSKLKGKLPNLYPPETIADGV	240	
Qy	241	KSSIGLNTWPIIRDLDVDDIFVTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFTQ	300	
Db	241	KSSIGLNTWPIIRDLDVDDIFVTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFTQ	300	
Qy	301	VSPEVKNICIVLSCGNVDLTSSITWVKAERPAASYQSVSV	340	
Db	301	VSPEVKNICIVLSCGNVDLTSSITWVKAERPAASYQSVSV	340	

```

1  RESULT 3
2  US-08-403-866-10
3  ; Sequence 10, Application US/08403866
4  ; Patent No. 5643779
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Ehrlich, Stanislaw
9  ; APPLICANT: Godon, Jean-Jacques
10 ; APPLICANT: Renault, Pierre
11 ;
12 ; TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate
13 ;
14 ; TITLE OF INVENTION: synthase from Lactococcus and its applications
15 ;
16 ; NUMBER OF SEQUENCES: 16
17 ;
18 ; COMPUTER READABLE FORM:
19 ;
20 ; MEDIUM TYPE: Floppy disk
21 ;
22 ; COMPUTER: IBM PC compatible
23 ;
24 ; OPERATING SYSTEM: PC-DOS/MS-DOS
25 ;
26 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
27 ;
28 ; CURRENT APPLICATION DATA:
29 ;
30 ; APPLICATION NUMBER: US/08/403,866
31 ;
32 ; FILING DATE:
33 ;
34 ; CLASSIFICATION: 435
35 ;
36 ; ATTORNEY/AGENT INFORMATION:

```

RESULT 4  
 US-08-669-378-2  
 ; Sequence 2, Application US/08669378  
 ; Patent No. 6107063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moeckel, Bettina  
 ; APPLICANT: Eggeing, Lothar  
 ; APPLICANT: Samh, Hermann  
 ; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
 ; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
 ; TITLE OF INVENTION: DEHYDRATASE  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-2

```

```

Query Match      17.9%; Score 310.5; DB 3; Length 436;
Best Local Similarity 27.8%; Pred. No. 1.4e-26;
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;

Qy 7 ISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRS 66
Db 21 IRAADIQAARISSVIAPTLPQYCPRLSEETGAETIYKREDLQDVRYSYKIRGALN---S 77
Qy 67 LVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVPOTAPDCCKKLAIOAYG--A 123
Db 78 GAOSPOQRDAGIVAASAGNHAOGVAVCKSLGVQGRYIVPVTPKOKRDRIMVHGFEV 137
Qy 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNTVIGQGTVAAEILSOLTSNGKSADH 197
Qy 180 LVVPVGGGMLAGIATVVKALKPSKVYAAEPSNADDCYOSKLGKLMPLNYPPEIADG 239
Db 198 VMVPPGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS--MQAALHNGGPIITLETVDPEVDG 256
Qy 240 VK--SSIGLNTWPIIRDLVDIFVTDEIKCATOL--VWERMKLLIEPTAGVGVAAVLSQ 296
Db 257 AEVQRVGDLYNTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGLIAEPAGALSIAGLKEM 316

```

```

RESULT 5
US-08-669-378-12
; Sequence 12, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGLATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-12

Query Match      17.9%; Score 310.5; DB 3; Length 436;
Best Local Similarity 28.1%; Pred. No. 1.4e-26;
Matches 94; Conservative 60; Mismatches 160; Indels 21; Gaps 9;

Qy 7 ISFADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRS 66
Db 21 IRAADIQAARISSVIAPTLPQYCPRLSEETGAETIYKREDLQDVRYSYKIRGALN---S 77
Qy 67 LVPDALEKPKAVVTHSSGNHGQALTYAAKLEGIPAVIVPOTAPDCCKKLAIOAYG--A 123
Db 78 GAOSPOQRDAGIVAASAGNHAOGVAVCKSLGVQGRYIVPVTPKOKRDRIMVHGFEV 137
Qy 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNTVIGQGTVAAEILSOLTSNGKSADH 197
Qy 180 LVVPVGGGMLAGIATVVKALKPSKVYAAEPSNADDCYOSKLGKLMPLNYPPEIADG 239
Db 198 VMVPPGGGGLLAGVSYVMADMAPRTAIVGIEPAGAAS--MQAALHNGGPIITLETVDPEVDG 256
Qy 240 VK--SSIGLNTWPIIRDLVDIFVTDEIKCATOL--VWERMKLLIEPTAGVGVAAVLSQ 296
Db 257 AEVQRVGDLYNTIVEKNQGRVHMMSATEGAVCTEMLDLYONEGLIAEPAGALSIAGLKEM 316

```

```

RESULT 6
US-08-669-378-4
; Sequence 4, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGLATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-4

Query Match 17.8%; Score 309.5; DB 3; Length 436;  
Best Local Similarity 27.8%; Pred. No. 1.9e-26;  
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;  
QY 7 ISPADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVR 66  
DB 21 IRAADIQTAAQTSSVIPTPLQYCPRLSEETGAEIYLKREDLDQVRSYKIRGALN---S 77  
QY 67 LVPDALKERPKAVVTHSSGNHGALTYAAKLEGIPAYIVVPTAPDCKKLAIOAYG---A 123  
DB 78 GAOSPOQRDAGIVAASAGNHAQGVAYVCKSLGVQGRYIVPVQTPKQKDRIMVHGGEFV 137  
QY 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNOVPLV---DA 179  
DB 138 SLVVTGNFDEASAAHEDARTGATLIEPFDARNVTIGOGTVAAILSLTSMGKSADH 197  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKMLPNLYPPETIADG 239  
DB 198 VMVPVGGGGLLAGVSVYMDAMPRTAIVGIEPAGAAS-MQAALHNGGPITLETVDVDFVDG 256  
QY 240 VK--SSIGLNTWPIIRDLVDDIFVTDEIKCATQL--VWERMKLLIETPAGVGVAAVLSQ 296  
DB 257 AEVKRVGDLNLTIVKNGQVRHMSATGAVCTEMLDLYQNEGIIAEPAGALSIAGLKEM 316  
QY 297 HFQTVSPVKNICIVLSSGGNVDLTSSITWVKQAR 331  
DB 317 SF---APGSVVVCI--SGGNNDV---LRYAEIAER 344

## RESULT 7

US-08-669-378-6  
Sequence 6, Application US/08669378.  
Patent No. 6107063  
GENERAL INFORMATION:  
APPLICANT: Moeckel, Bettina  
APPLICANT: Eggeling, Lothar

APPLICANT: Sahn, Hermann  
TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF  
TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE  
TITLE OF INVENTION: DEHYDRATASE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,378  
FILING DATE: 20-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE95/00017  
FILING DATE: 09-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 00 926.7  
FILING DATE: 14-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016881/0142  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-669-378-6

Query Match 17.8%; Score 309.5; DB 3; Length 436;  
Best Local Similarity 27.8%; Pred. No. 1.9e-26;  
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;  
QY 7 ISPADVEKAHINIRDSIHLTPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVR 66  
DB 21 IRAADIQTAAQTSSVIPTPLQYCPRLSEETGAEIYLKREDLDQVRSYKIRGALN---S 77  
QY 67 LVPDALKERPKAVVTHSSGNHGALTYAAKLEGIPAYIVVPTAPDCKKLAIOAYG---A 123  
DB 78 GAOSPOQRDAGIVAASAGNHAQGVAYVCKSLGVQGRYIVPVQTPKQKDRIMVHGGEFV 137  
QY 124 SIYVCEPSDESRENKRVTEETEGIMVHPNQEPVAVIAGOGTIALEVLNOVPLV---DA 179  
DB 138 SLVVTGNFDEASAAHEDARTGATLIEPFDARNVTIGOGTVAAILSLTSMGKSADH 197  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKMLPNLYPPETIADG 239  
DB 198 VMVPVGGGGLLAGVSVYMDAMPRTAIVGIEPAGAAS-MQAALHNGGPITLETVDVDFVDG 256  
QY 240 VK--SSIGLNTWPIIRDLVDDIFVTDEIKCATQL--VWERMKLLIETPAGVGVAAVLSQ 296  
DB 257 AEVKRVGDLNLTIVKNGQVRHMSATGAVCTEMLDLYQNEGIIAEPAGALSIAGLKEM 316  
QY 297 HFQTVSPVKNICIVLSSGGNVDLTSSITWVKQAR 331  
DB 317 SF---APGSVVVCI--SGGNNDV---LRYAEIAER 344

## RESULT 8

```
US-08-669-378-10
; Sequence 10, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-10

Query Match 17.8%; Score 309.5; DB 3; Length 436;
Best Local Similarity 27.8%; Pred. No. 1.9e-26;
Matches 93; Conservative 61; Mismatches 160; Indels 21; Gaps 9;

QY 7 ISFADVEKAHINTRDSIHLPVLTSSILNOLTGRLFFKCELFQKGTGSKIRGALNAVRS 66
Db 21 IRAADIOTAQARISSVIAPTLOQYCPRLSEETGAETYLKREDLQDVRSYKIRGALN---S 77
QY 67 LVPDALERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPDCKKLAIOAYG---A 123
Db 78 GAQSPQEQDAGIVAASAGNHAOGVAYVCKSLGVQGRIVVPTQPKOKRDRIMVHGFEV 137
QY 124 SIYCEPSPDESRENKRVTEETEGIMVHPNOEPVIAQGQTTALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNVIGOGTVAAEILSOLTSMGKSADH 197
QY 180 LVVPGVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKMLPNLPYPPETIADG 239
Db 198 VVMPVPGGGGLAGVSYWMDMAPRTAIVGTIEPAGAAS--MQAALHNGGPITLTETVDFVDG 256
QY 240 VK--SSGLNTWPIIRDVDDIFVTVEDEIKCATOL--VWERMKLLIETAGVCAVAVLSQ 296
Db 257 AEVKRVGDLNLTIVTEKNOGRVHMMSATEGAVCTEMLDLYONEGIIAEPAAGALSIAKLEM 316

US-08-669-378-8
; Sequence 8, Application US/08669378
; Patent No. 6107063
; GENERAL INFORMATION:
; APPLICANT: Moeckel, Bettina
; APPLICANT: Eggeling, Lothar
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
; TITLE OF INVENTION: RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
; TITLE OF INVENTION: DEHYDRATASE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,378
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE95/00017
; FILING DATE: 09-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 00 926.7
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016881/0142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-669-378-8

Query Match 17.7%; Score 307.5; DB 3; Length 436;
Best Local Similarity 28.6%; Pred. No. 3.1e-26;
Matches 96; Conservative 59; Mismatches 158; Indels 23; Gaps 10;

QY 7 ISFADVEKAHINTRDSIHLPVLTSSILNOLTGRLFFKCELFQKGTGSKIRGALNAVRS 66
Db 21 IRAADIOTAQARISSVIAPTLOQYCPRLSEETGAETYLKREDLQDVRSYKIRGALN---S 77
QY 67 LVPDALERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTQAPDCKKLAIOAYG---A 123
Db 78 GAQSPQEQDAGIVAASAGNHAOGVAYVCKSLGVQGRIVVPTQPKOKRDRIMVHGFEV 137
QY 124 SIYCEPSPDESRENKRVTEETEGIMVHPNOEPVIAQGQTTALEVLNQVPLV----DA 179
Db 138 SLVVTGNFDEASAAHEDAERTGATLIEPFDARNVIGOGTVAAEILSOLTSMGKSADH 197
QY 180 LVVPGVGGGMLAGIATTVKALPKSVKYAAEPSNADDCYQSKLKGKMLPNLPYPPETIADG 239
Db 198 VVMPVPGGGGLAGVSYWMDMAPRTAIVGTIEPAGAAS--MQAALHNGGPITLTETVDFVDG 256
QY 240 VK--SSGLNTWPIIRDVDDIFVTVEDEIKCATOL--VWERMKLLIETAGVCAVAVLSQ 296
Db 257 AEVKRVGDLNLTIVTEKNOGRVHMMSATEGAVCTEMLDLYONEGIIAEPAAGALSIAKLEM 316
```

Db 198 VMPVPGGGGLLAGVWSYMDAMPRTAIVGTEPAGAAS-MQAAJHNGSPITLETVPFVDG 256  
QY 240 --VKSSIGLNTWPIIRDVLVDIFVTVEDEIKCATQL--VWERMKLLIEPTAGVGAVALS 295  
Db 257 GEYKRVGDLN-YTIVEKNQGRVHMSTEGAVCTEMLDLYONEGIIAEPAGALSIAGLKE 315  
QY 296 QHFQTVSPVKNKICIVLSGGNVDLTSSITVWQAE 331  
Db 316 MSF---APGSVWVCII-SGGNDV---LRYAEIAER 344

RESULT 10  
US-09-789-300A-4  
; Sequence 4, Application US/09789300A  
; Patent No. 6458576  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 22406, A No. 6458576el Human Pyridoxal-Phosphate  
; FILE REFERENCE: 35800/208926  
; CURRENT APPLICATION NUMBER: US/09/789,300A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: US 60/183,208  
; PRIOR FILING DATE: 2000-02-17  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; NUMBER OF SEQ ID NOS: 4  
; SEQ ID NO 4  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Pyridoxal-Phosphate Dependent Enzyme Family Domain  
; OTHER INFORMATION: Sequence  
US-09-789-300A-4

Query Match 16.3%; Score 283.5; DB 4; Length 378;  
Best Local Similarity 28.3%; Pred. No. 1.4e-23;  
Matches 109; Conservative 55; Mismatches 126; Indels 95; Gaps 15;

QY 19 IRDSIHLPVLTSSILNQLTGRNL-----PFKCE-LFQKTSFKIRG-ALNAV---- 64  
Db 1 VTELGNTPLVRNLRUSKELGEGLGANAAVEIYKLEDLNGPTGSPKDRGLALNMTLLAE 60  
QY 65 -----RSLVPDALERPK-AVTHSSGNHQALTYAAKLEGIPAYIVWPOTAPDCKKLA- 117  
Db 61 KLGGKGGIVPGTVQVESKTIIEPTSGNTGIALALAAALLGLKCTIVPATDTSREKRAQ 120  
QY 118 IQAYGASIVCEPSDESRENKRVTEETEGIMVHPNQEPAV-----TA 161  
Db 121 LRALGAELVVPVAGGSDDLADAIKAE--LAEENPENAYLLNQAAGPFONPANPEIA 178  
QY 162 GQCTIALELVNQV-----PLVDALVVPVGGGMLAGIATVVKALPS-----V 204  
Db 179 GQKTIGPEIWEQGGKEISLGRLPDAVAVPVGGGTITGIARYLKEINPDGKIDVLELVP 238  
QY 205 KYVAEPPSNADDCYOSKLAGKLMPLY-----PPTIADGVKSSIG 245  
Db 239 KVIGVEPGS-----AVLSGLSKATITLAGKCPHLGRDSKYLLODEPTVLTETKSIG 293  
QY 246 LN-----TWPIIRD-----VDDIFTVEDEIKCATQLVWERMKLLIEPTAGVGA 292  
Db 294 LGVPRVGEFVPPIDELDRRQIDEVTVTDEEALEAARLLAREGILVGPSSGA 353  
QY 293 VLSQHPQTVSP--EVKNICIVLSGG 315  
Db 354 ALKLAKEGKPLNKGKTIIVILSGG 378

RESULT 11

US-09-088-435-1

; Sequence 1, Application US/09088435

; Patent No. 6277619

; GENERAL INFORMATION:  
; APPLICANT: LAL, PREETI  
; APPLICANT: CORLEY, NEIL C.  
; APPLICANT: GUEGLER, KARL J.  
; APPLICANT: PATTERSON, CHANDRA  
; TITLE OF INVENTION: SERINE DEHYDRATASE HOMOLOG  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/088,435  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cerrone, Michael C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0512 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: TPIAZS08  
; CLONE: 2752518  
; US-09-088-435-1

Query Match 15.0%; Score 259.5; DB 4; Length 325;

Best Local Similarity 28.6%; Pred. No. 5.8e-21;

Matches 97; Conservative 58; Mismatches 113; Indels 71; Gaps 16;

QY 10 ADVERAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRLVP 69

Db 9 AKQEPFHV-----VTPLESWALSQVAGMPVLKCNVQPSGSKIRG-----IGHFCQ 57

QY 70 DALEKRPKAVVTHSSGNHQALTYAAKLEGIPAYIVWPOTAPDCKKLAIQAYGAST---- 125

Db 58 ENAKGCRHLVCSGSGNAGIAAAYAAKLGIPATIVLPSTSLQVVQRLOGEGAEVQLRG 117

QY 126 -VYCEPSDESRENKRVTEETEGIMVHPNQEPVIAQGTIALE---VLNQVPLVDALV 181

Db 118 KYWDEANLRAQE-LAKRDGWN-----VPPFDHPLIMKHASLVQELKAVLRTPP--GALV 170

QY 182 VPWGGGMLAGIATVVKAL-----KPSVKVYAAEPSNADDCYOSKL-KGLMPNLYPETI 236

Db 171 LAVGGGGLAGVAGLVLEGVQHOVPIIAHGAH-----CFNAAATAGKLV-----TL 217

QY 237 AD--GVKSSIGLNT-----WPIIRDVLVDIFVTVEDEIKCATQLVWERMKLLIE 283

Db 218 PDITSVAKSLGAKTVAARALECMQVCKTHSEVEDTEAVS-----AVQQLDDERMLVE 271

QY 284 PTAGVGAVALS-----QHFQTVSPVKNKICIVLSGGN 316

Db 272 PAC--GAAIYSGLLRRLRQAEGCLPPLSLTSVVVIVCGGN 308

RESULT 12  
US-09-134-001C-4168  
; Sequence 4168, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4168  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4168

Query Match 13.9%; Score 241.5; DB 4; Length 367;  
Best Local Similarity 28.1%; Pred. No. 8.1e-19;  
Matches 93; Conservative 58; Mismatches 149; Indels 31; Gaps 12;  
QY 26 TPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAVRSVLPDALKRPAVVTHSSG 85  
DB 44 TPLIYCNLSOKLDIELYKYGANPTGSKRGVMVATK----AKEGKKVVICASTG 99  
QY 86 N-HGQALTYAAKLEGIPAVIVVPTAPDCKKLAIAQA--YGASIVYCEPSDESRENVAKR 142  
DB 100 NTSASAAVAAAR-AGLKATVWPEGKIALGKLS-QAVWYGAIEVSTEGNDEALEIVKEI 157  
QY 143 TEETEGIMVHPNOEPAVIAQGTIALEVLNQV--FLVDALVVPVGGGMLAG-----TAI 195  
DB 158 AEENDEIELVSNVPRIEGQKTAPEIVEQLDGOAPDILATPVGNAGNITAYWQGFVEY 217  
QY 196 TVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGVKSSIG-LNTWPIIRD 254  
DB 218 HNKNTQLPOMGFQOEGASPIVQNKI-----TKNPETIATAIR--IGNPASQKAVN 268  
QY 255 LVDD----IFTVTEDEIKATOLVWERMKLLIEPTAGVGAVALVSOHQFQTVSPVKNICI 310  
DB 269 ALDESGLIDSVIDEILEAYQLMTTNEGVSFEPASNASIAGLIKLRSGKLPKCKKIWA 328  
QY 311 VLSG-GNVLTSSITWVQKARPASVQSVSV 340  
DB 329 ILTGNGLKDPDTAISLLDNPIQLPNNKESI 359

RESULT 13  
US-08-120-960-2  
; Sequence 2, Application US/08120960  
; Patent No. 5523225  
; GENERAL INFORMATION:  
; APPLICANT: KRAUS, JAN P  
; TITLE OF INVENTION: DNA SEQUENCE ENCODING HUMAN  
; TITLE OF INVENTION: CYSTATHIONINE B-SYNTHASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: DILWORTH & BARRESE  
; CITY: SAN DIEGO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APLICATION DATA:  
; APPLICATION NUMBER: US/08/120,960  
; FILING DATE: 12-SEP-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PEPPER PH.D., FREDERICK W.  
; REGISTRATION NUMBER: 31,286  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4410  
; TELEFAX: 619-453-2839  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-120-960-2  
Query Match 10.8%; Score 187; DB 1; Length 551;  
Best Local Similarity 25.2%; Pred. No. 2.7e-12;  
Matches 86; Conservative 46; Mismatches 149; Indels 60; Gaps 11;

QY 26 TPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAVRSVLPDALKRPAVVTHS 83  
DB 87 TPMVRINKTKGKFKGLKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGLKPGDTIIEPT 146  
QY 84 SCNHQALTYAAKLEGIPAVIVVPTAPDCKKLAIAQYASIVYCE-----PSDESREN 138  
DB 147 SONTGIGLALAAVARGYRCIIIVPEKMSSEKVDVLRALGAEIVRTNARFDSPESHVGV 206  
QY 139 AKRVTTEE--TEGIMVHPNOEPAVIAQGTIALEVLNQV- LVDAVVPVGGGMLAGIAI 195  
DB 207 AWRLKNEIPNSHILDOYRNASNP LAHYDTTAD EILQOCDCGKLDMLVASVGTGTTIGIAR 266  
QY 196 TVKALKPSVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGVKSS-----IGLNTWP 250  
DB 267 KLKCKPCGCRITGVDP-----EGSI---LAEPEELNQTQETTYEVEGIGYDFIP 312  
QY 251 II--RDVDDIFTVTEDEIKATOLVWERMKLLIEPTAG--VGVAAVLVSOHQFQTVSPVK 306  
DB 313 TVLDRTVVDKFKSNDEEAFTFARMLIAQOGLGGSAGSTVAVAYAKAQLQE-----G 367  
QY 307 NTCIVLSGQNV-----DLTSSITW 325  
DB 368 QRCVVILPDSVRNYMTKFLSDRWMLQKGLKEDLTKKPKW 408

RESULT 14  
US-09-347-878-9  
; Sequence 9, Application US/09347878C  
; Patent No. 6376210  
; GENERAL INFORMATION:  
; APPLICANT: Yuan, Chong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES  
; FILE REFERENCE: 25885-1651  
; CURRENT APPLICATION NUMBER: US/09/347,878C  
; CURRENT FILING DATE: 1999-07-06  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-347-878-9

Query Match 10.8%; Score 187; DB 4; Length 551;  
Best Local Similarity 25.2%; Pred. No. 2.7e-12;  
Matches 86; Conservative 46; Mismatches 149; Indels 60; Gaps 11;  
QY 26 TPVLTSSILNQLTGR--NLFFKCELFQKTSFKIRGALNAVRSVLPDALKRPAVVTHS 83  
DB 87 TPMVRINKTKGKFKGLKCELLAKCEFFNAGGSVKDRISLRMIEDAERDGLKPGDTIIEPT 146

Job time : 24.5317 secs

```
QY      84 SGNHGQALTYAAKLEGIPAYIVVQTPADCKKLAIQAYGASIVYCE-----PSDESRENV 138
Db      147 SGTGIGLALAAAVRGYRCIIIVPEKMSSEKVDVLRALGAIEVTRPTNARFDSPESHVGV 206
QY      139 AKRVTEE--TEGIMVHPNOEPAVIAQGQITALEVLNOVP-LYDALVVPVGGGMLAGIAI 195
Db      207 AWRLKNEINSHILDYRNASPLAHYDTTAEIIOQCDKLDMLVASVGTGGTITGIAR 266
QY      196 TVKALPSPVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADGVKSS-----IGLNTWP 250
Db      267 KLKEKCPGCGRIIGVDP-----EGSI---LAEPEELNQTGTTVEVIGYDFIP 312
QY      251 II--RDLYDDITVTVEDEIKCATQLWERMKLLIEPTAG--VGVAAVLVSOHFQTVSPYVK 306
Db      313 TVLDRVTVDKPKSNDDEEAFTEARMLIAOEGLLCGGSAGSTVAVAKAAQELQE-----G 367
QY      307 NICIVLSGGNV-----DLTSSITW 325
Db      368 QRCVILPDSVRNRYMTKFLSDRWMLQKGPLKBEEDLTEKRPW 408

RESULT 15
US-09-134-001C-3920
; Sequence 3920 Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3920
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3920

Query Match      10.0%; Score 174; DB 4; Length 312;
Best Local Similarity 23.6%; Pred. No. 3.1e-11;
Matches 74; Conservative 48; Mismatches 128; Indels 64; Gaps 11;

QY      21 DSIHLTPVLTSILNQLTGRN--LFFKCELFQKTSFKIRGALNAVRSLVDPALERKPKA 78
Db      15 DLIGQTPLY---LLESFSDENKIYAKLEQFNPGGSIKDRLGKYLIEKAIDEGRIKEGDT 71
QY      79 VVTHSSGNHGQALTYAAKLEGIPAYIVVQTPADCKKLAIQAYGASIVYCEPSDESRENV 138
Db      72 IVETAGNTGIGLATASNRHKVKCIIFAPGFAEEKISIMKALGADV-----118
QY      139 AKRVTEETEGIMVHPNOEPAVIAQGQ-----GTIALEVLNOVPLVDA 179
Db      119 --RRTPKAEG--WTGAQOEALAVATRYGLYLMNQFETKDNPGAYTQTLAKQLTDELSDHY 175
QY      180 LVVPVGGGMLAGIAITVYKALPSPVKVYAAEPSNADDCYQSKLKGKLPNLYPPETIADG 239
Db      176 FVAGVSGGTTGTGVAQHLKTY--DVKNYIVEPEG-----SVLNGGVS---HPHAT---220
QY      240 VKSSIGLNTWP--IIRDLVDDITVTVEDEIKCATQLWERMKLLIEPTAGVGVAAVLSQH 297
Db      221 --EGIGSEKWPSELEKELVDGIFTVADKDAFNVKLVANKEGLLVGSSSGAALQGALE---276
QY      298 FQTVSPYKNCIV 311
Db      277 ---LKKSIQNGVIV 287
```

Search completed: June 24, 2003, 05:36:46



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:11:01 ; Search time 276.907 Seconds  
(without alignments)  
132.862 Million cell updates/sec

Title: us-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYICISFADVEKAHINR.....SSITWVQARFPASYOSVSU 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	349.5	20.1	310	9	US-09-738-626-4591 Sequence 4591, Ap
2	311.5	18.0	436	9	US-09-738-626-5828 Sequence 5828, Ap
3	192.5	11.1	306	10	US-09-815-242-13657 Sequence 13657, A
4	190.5	11.0	308	9	US-09-738-626-6317 Sequence 6317, Ap
5	174.5	10.1	310	10	US-09-815-242-10707 Sequence 10707, A
6	169.5	9.8	313	10	US-09-815-242-12146 Sequence 12146, A
7	169.5	9.8	313	10	US-09-815-242-12800 Sequence 12800, A
8	169.5	9.8	323	10	US-09-815-242-10239 Sequence 10239, A
9	167.5	9.7	303	10	US-09-815-242-5437 Sequence 5437, Ap
10	161.5	9.3	323	10	US-09-815-242-13862 Sequence 13862, A
11	154	8.9	383	9	US-09-931-457A-65 Sequence 65, Appl
12	149	8.6	374	10	US-09-815-242-5737 Sequence 5737, Ap
13	148	8.5	325	9	US-09-931-457A-32 Sequence 32, Appl
14	147	8.5	324	9	US-09-905-290A-4 Sequence 4, Appl
15	147	8.5	386	9	US-09-931-457A-66 Sequence 66, Appl
16	147	8.5	398	9	US-09-931-457A-62 Sequence 62, Appl
17	146	8.4	415	9	US-09-931-457A-64 Sequence 64, Appl
18	143.5	8.3	326	10	US-09-815-242-11179 Sequence 11179, A
19	143.5	8.3	324	10	US-09-815-242-11879 Sequence 11879, A

20	139	8.0	325	9	US-09-931-457A-31 Sequence 31, Appl
21	131	7.6	306	10	US-09-815-242-11292 Sequence 11292, A
22	120	6.9	317	9	US-09-738-626-5838 Sequence 5838, Ap
23	115.5	6.7	250	10	US-09-815-242-11635 Sequence 11635, A
24	112	6.5	1187	10	US-09-935-291A-4 Sequence 4, Appl
25	105	6.1	540	9	US-09-742-096-22 Sequence 22, Appl
26	105	6.1	1786	9	US-09-742-096-3 Sequence 3, Appl
27	101.5	5.9	212	9	US-09-742-096-25 Sequence 25, Appl
28	101.5	5.9	630	9	US-09-742-096-5 Sequence 5, Appl
29	99	5.7	1642	10	US-09-925-442-2 Sequence 2, Appl
30	96.5	5.6	26926	9	US-09-759-508B-2 Sequence 2, Appl
31	95.5	5.5	472	10	US-09-745-763-36 Sequence 36, Appl
32	95.5	5.5	473	9	US-09-984-271-120 Sequence 120, App
33	95	5.5	164	10	US-09-734-017A-28 Sequence 28, Appl
34	95	5.5	577	10	US-09-815-242-13870 Sequence 13870, A
35	93.5	5.4	852	10	US-09-815-242-11535 Sequence 11535, A
36	93	5.4	577	10	US-09-741-669-354 Sequence 354, App
37	93	5.4	577	10	US-09-815-242-10286 Sequence 10286, A
38	93	5.4	1648	10	US-09-925-442-35 Sequence 35, Appl
39	91.5	5.3	833	9	US-09-928-457-40 Sequence 40, Appl
40	90.5	5.2	442	9	US-10-117-417-6 Sequence 6, Appl
41	90.5	5.2	442	9	US-10-117-417-16 Sequence 16, Appl
42	89.5	5.2	853	10	US-09-815-242-11363 Sequence 11363, A
43	89	5.1	815	9	US-10-198-070-123 Sequence 123, App
44	87.5	5.0	302	9	US-09-738-626-6862 Sequence 6862, Ap
45	87.5	5.0	319	10	US-09-815-242-13897 Sequence 13897, A

ALIGNMENTS

RESULT 1

US-09-738-626-4591  
; Sequence 4591, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/097738.626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4591  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4591

Query Match 20.1%; Score 349.5; DB 9; Length 310;  
Best Local Similarity 29.1%; Pred. No. 8.2e-25;  
Matches 95; Conservative 55; Mismatches 146; Indels 31; Gaps 7;

Qy	7	ISFADVEKAHINIRDSIHLTPVLTSSILNLTGNLFKCELFKGTGSKTRGALNAVR	66
Db	2	LTLDVITAQORAPHRVTRTLPFEA---DPIDGTQIWKIAEFLKQCGVKTRGAFNQLA	58
Qy	67	LVPDALEKPKAVVTHSSNGHQAITYAAKLEGIPAYIVVPTQAPDCKKLAIQAYGASIV	126





; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12146  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12146

Query Match 9.8%; Score 169.5; DB 10; Length 313;  
Best Local Similarity 24.5%; Pred. No. 8.4e-08;  
Matches 74; Conservative 50; Mismatches 143; Indels 35; Gaps 10;  
  
QY 9 FADVEKAHINIRDSIHLTPVLSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSVLV 68  
Db 2 FLMAQKPVNDITQIIIGTTPVKLRNVDDNAAADVVKLEVQNPGGSVKORIALAMTEKAE 61  
  
QY 69 PDALERKPKAVVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPDCKKLAIQAGASIVYC 128  
Db 62 REGKIRPGDTIVEPTSGNTGIGLAFVCAAKGYKAVFTMPETMSOERNLLKAYGAEVLV-L 120  
  
QY 129 EPSDESRENVAKRVTE--ETEGIMVHPNO-----EPAVIAGOGTTALEVLNV--PLVDA 179  
Db 121 TPGSEAMKAIGAKKELKEEHGYF-EPQOFENPANPEV--HELTGPELLOQFEGKTIDA 177  
  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPS-----NADDCYOSKLGKMLPNLYPPE 234  
Db 178 FLAGVGTGCTLSGVKVLKKEYPNTEIVAIEPEASVLSGGEPGPHKLOG--LGAGFIPG 235  
  
QY 235 TIADGVKSSIGLNTWPIIRDLDVDDIFTVTEDEIKCATQLVWERMKLLIPTAGVGVAAVL 294  
Db 236 T-----LNT-----EYDSIIKVGNDTAMEMSRVAKREEGILAGISSGAATYAAI 280  
  
QY 295 SQ 296  
Db 281 QK 282

RESULT 7  
US-09-815-242-12800  
; Sequence 12800, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12800  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12800  
  
Query Match 9.8%; Score 169.5; DB 10; Length 313;  
Best Local Similarity 24.5%; Pred. No. 8.4e-08;  
Matches 74; Conservative 50; Mismatches 143; Indels 35; Gaps 10;  
  
QY 9 FADVEKAHINIRDSIHLTPVLSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSVLV 68  
Db 2 FLMAQKPVNDITQIIIGTTPVKLRNVDDNAAADVVKLEVQNPGGSVKORIALAMTEKAE 61  
  
QY 69 PDALERKPKAVVTHSSGNHGOALTYAAKLEGIPAYIVVPTAPDCKKLAIQAGASIVYC 128  
Db 62 REGKIRPGDTIVEPTSGNTGIGLAFVCAAKGYKAVFTMPETMSOERNLLKAYGAEVLV-L 120  
  
QY 129 EPSDESRENVAKRVTE--ETEGIMVHPNO-----EPAVIAGOGTTALEVLNV--PLVDA 179  
Db 121 TPGSEAMKAIGAKKELKEEHGYF-EPQOFENPANPEV--HELTGPELLOQFEGKTIDA 177  
  
QY 180 LVVPVGGGMLAGIATTVKALPKSVKYAAEPS-----NADDCYOSKLGKMLPNLYPPE 234  
Db 178 FLAGVGTGCTLSGVKVLKKEYPNTEIVAIEPEASVLSGGEPGPHKLOG--LGAGFIPG 235  
  
QY 235 TIADGVKSSIGLNTWPIIRDLDVDDIFTVTEDEIKCATQLVWERMKLLIPTAGVGVAAVL 294  
Db 236 T-----LNT-----EYDSIIKVGNDTAMEMSRVAKREEGILAGISSGAATYAAI 280  
  
QY 295 SQ 296  
Db 281 QK 282

RESULT 8  
US-09-815-242-10239  
; Sequence 10239, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

Query Match	9.7%;	Score 167.5;	DB 10;	Length 303;
Best Local Similarity	24.6%;	Pred. No. 1.2e-07;		
Matches	72: Conservative	49: Mismatches	137: Indels	35: Gaps

Qy	18	NIRDSIHLTPVLTSIIINOLTGRLNLFKCELFQKGTSGFKIRGALNAVRSRLVPDALERPK	77
Db	2	NITQIIIGTTPVKILURNVDDNAADVVYKLEYONFGSVKDRIALAMIEKAEREGRKPGD	61
Qy	78	AVVTHSSGNHQALTYAAKLEGIPAYIVVPTADCKKLAIQAYGASIVCEPSPDESREN	137
Db	62	TIVETPSGNTGIGLAFCAAKGYKAVFTMPETMSOERNLLIKAYGAELY-LTPGSEAMKG	120
Qy	138	VAKRVTE--ETEGIVWHPNQ-----EPAVIAQGTIALEVINQV--PLVDVALVVPVGGG	188
Db	121	AIKKAKELKEBHGXF-EPQOPENPANPEV--HELTGPELLQQFEGKIDAFLAGVGTGG	177
Qy	189	MLAGIAITVAKLKPYSVKYAAEPS-----NADDCCYQSKLCKLMPNLNPPETIADGVSKS	243
Db	178	TLSCGVKVLKEYENIEIVALEPASPVLSGEPGPHKQQ--LGAAGFIPT-----	227
Qy	244	IGLNTWPIIRDLVDDITFTVDETEKCATQLVWERMKLLIETAGVGAVALISQ	296
Db	228	--LNT-----EIVDSIIKVGNDTAMENSRVRKAEIGILAGISSGAIIYAAAIQK	273

```

RESULT 10
US-09-815-242-13862
; Sequence 13862, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EUTRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13862
; LENGTH: 323
; TYPE: prt
; ORGANISM: Salmonella typhi

```

```

> APPLICANT: Zyskind, Judith W.
> APPLICANT: Wall, Daniel
> APPLICANT: Trawick, John D.
> APPLICANT: Carr, Grant J.
> APPLICANT: Yamamoto, Robert T.
> APPLICANT: Xu, H. Howard
> TITLE OF INVENTION: Identification of Essential Genes in
> TITLE OF INVENTION: Prokaryotes
> FILE REFERENCE: ELITRA.011A
> CURRENT APPLICATION NUMBER: US/09/815,242
> CURRENT FILING DATE: 2001-03-21
> PRIOR APPLICATION NUMBER: 60/191,078
> PRIOR FILING DATE: 2000-03-21
> PRIOR APPLICATION NUMBER: 60/206,848
> PRIOR FILING DATE: 2000-05-23
> PRIOR APPLICATION NUMBER: 60/207,727
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: 60/242,578
> PRIOR FILING DATE: 2000-10-23
> PRIOR APPLICATION NUMBER: 60/253,625
> PRIOR FILING DATE: 2000-11-27
> PRIOR APPLICATION NUMBER: 60/257,931
> PRIOR FILING DATE: 2000-12-22
> PRIOR APPLICATION NUMBER: 60/269,308
> PRIOR FILING DATE: 2001-02-16
> NUMBER OF SEQ ID NOS: 14110
> SOFTWARE: FastSEQ for Windows Version 4.0
> SEQ ID NO 13862
> LENGTH: 323
> TYPE: prt
> ORGANISM: Salmomella typhi

```

```

; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(323)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13862

Query Match
Best Local Similarity 9.3%; Score 161.5; DB 10; Length 323;
Best Local Similarity 23.0%; Pred. No. 5e-07;
Matches 76; Conservative 63; Mismatches 126; Indels 65; Gaps 14;

QY 26 TPVLTSSILNQLTGRNLFKCELPQKTSFKIRGALNAVRLVDPDALERKPKA-VVT 81
Db 15 TPLVR---LNRIGNGRILAKVE--SRNPSESVKCRIGA--NMIWDAEKRGVLKPGVBLE 67
QY 82 HSSGNHGOALTYAAKLEGIPAYIVVPOTPADCKKLAIOAYGASIVYCEPS-----DES 134
Db 68 PTSNGTIALAYNAARGYKLTLPMPETMSIERKKLKGALNVLTEGAGKMGKAIOKA 127
QY 135 RENYAK-----RVTEETEGIMVHPNOEPAVIAQGTIALEVLNQVPLV 177
Db 128 EEIVASDPQKYLQLQSFSPANPEIHEKTTGPEIWDTD-----GQ-----V 169
QY 178 DALVVPVGGGMLAGIAITVAKLPSVKV--YAAEPSNADDCYQSKLKGKMLNLPPEP 235
Db 170 DVFISGVTGGLTVTRYIKGTGKGTDLITVAVEPTDSPVIAQALAGEBIKPGPHKIQ 229
QY 236 IADG-VKSSIGLNTWPIIRLDVDDIFTVTDEIKCATOLVWERMKLLIEPTAGVGAANL 294
Db 230 IGAGFIPGNLDLK-----LIDKVGITWEEALSTARRLWEEGILAGISSGAANAAL 282
QY 295 SQHFQTVSPVKNICIVL-SGGNVDLTSSI 323
Db 283 KIQ-EDSEFTKNVILPSSGERYLSTAL 311

RESULT 11
US-09-931-457A-65
; Sequence 65, Application US/09931457A
; Patent No. US20020157132A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes
; FILE REFERENCE: BB1116 US CIP
; CURRENT APPLICATION NUMBER: US/09/931.457A
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 09/424,976
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/065,385
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/049,406
; PRIOR FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 65
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Spinacia oleracea
US-09-931-457A-65

Query Match
Best Local Similarity 8.9%; Score 154; DB 9; Length 383;
Best Local Similarity 22.9%; Pred. No. 3.3e-06;
Matches 75; Conservative 59; Mismatches 130; Indels 64; Gaps 15;

QY 6 CISFADVEKAHINRISIHLPVLTSSILN-----QLTGR-----NLFF 44
Db 38 CNNFKVVSSPITCK-AVLSPPSTIEGLNIAEDVSLGKTPMVLNNVSKGSVANIAA 96
QY 45 KCFLFQKTSFKIRGALNAVRLVDPDALERKPKAVVT-----HSSGNHGOALTYAAK 96
Db 97 KLESMEPCSVKDRIGY-----SMIDDA---EQKVITPCKTTLVETSGNTGIGLAFIA 149
QY 97 LEGIPAYIVVPOTPADCKKLAIOAYGASIVYCEPSD-----ESRENVAKRVTEETEGIM 150

; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(323)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5737

Query Match
Best Local Similarity 8.6%; Score 149; DB 10; Length 314;
Best Local Similarity 21.9%; Pred. No. 7.3e-06;
Matches 77; Conservative 62; Mismatches 144; Indels 68; Gaps 12;

QY 18 NTRDSTHLTPVLTSSILNQLTGRN-LFFKCELPQKTSFKIRGALNAVRLVDPDALERK 76
Db 2 SLSDSVGQTPMVQ---LHQLFPKHEVFAKLEVYMNPGSKMRDPKAKYIIIEHGIKHGLITEN 58
QY 77 KAVVTSSNGHGOALTYAAKLEGIPAYIVV-POTAPDCKKLAIOAYGASIVYCEPSDES 135
Db 59 THLIESTNGGALAMIAIKGLKLTCTVDPKRISPTNLKI-IKSYGANVMVVEEDPAHG 117
QY 136 ENVAKRVTEETE-----GIMVHPNQ--EPAVIAGQGTIALEVLNQVPLVDALVVPV 184
```

Db 118 GYLMTRIKQVQELLATIDDAYWYNQYANELNMOSHYGAGTEIVETIKQP--IDYEVAPV 175  
QY 185 GGGGMLAGTAITVKALKSVKVAAPSNADDCYQSKGLKGLMPNLYPPETIADGVKSSI 244  
Db 176 STTSGSINGMRKIKEVHPNQAIIAVD-----AKGSV 206  
QY 245 GLNWTPIIRD-----VDDIFTVTDEIK-----CATOLVWERMKLLIETAGVG 289  
Db 207 IFGDKPINRELPGIGASVRPNEXIRNXSSDPCRXLSICFGLSKTDXLXRHICGGSTGS 266  
QY 290 VAAVLSQHFQTSPEVKNICIVLSGN--VDLTSSITWYKQABRPASVQSV 338  
Db 267 IIAAIEQLITSIEGATIVTILDRGDRYLDLVYSDTWL---EKMRSRGV 314

## RESULT 13

US-09-931-457A-32

; Sequence 32, Application US/09931457A

; Patent No. US20020157132A1

; GENERAL INFORMATION:

; APPLICANT: Falco, S. Carl

; APPLICANT: Allen, Stephen M.

; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

; FILE REFERENCE: BB1116 US CIP

; CURRENT APPLICATION NUMBER: US/09/931,457A

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: 09/424,976

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/065,385

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/049,406

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 32

; LENGTH: 325

; TYPE: PRT

; ORGANISM: Citrullus lanatus

US-09-931-457A-32

Query Match 8.5%; Score 148; DB 9; Length 325;  
Best Local Similarity 22.3%; Pred. No. 9.5e-06;  
Matches 64; Conservative 45; Mismatches 98; Indels 80; Gaps 10;

QY 45 KCELFQKTSFKIRGALNAVRSLVDPDALEK-----PKAVVTHSSGNHQALTYAAKLE 98  
Db 38 KLEMEPCSVKDRIGY----SWISDA-ENKGLITPCESVLIETPGSGTIGLAFIAAAK 92  
QY 99 GIPAYIVVPQTAPDCKKLAIQAYGASIVYCEPS-----DESRENVAK-----140  
Db 93 GYRLIICMPASMSLERRTILRAFGAELVLPDARGMKGAVOKAEIKAKTPNSYILQQFE 152  
QY 141 -----RVTEETEGIMVHPNOEPAVIAGQGTIALEVLNQPVLVDALVVPVGGGMLAGIAI 195  
Db 153 NPANPKIHYETT-----PEIWRGSG-----GKIDALVSGIGTGTVTCAGK 194  
QY 196 TVKALKPSVKVYAAEPSNADDCYQSKLKGKLMNLYPPETIADGVK-----SSIGLNTW 249  
Db 195 YLKEQNPNIKLYGVEPES-----AIISSGGPGPHKIOGIGAGFI 234  
QY 250 PIIRD--LVDDIFTVTDEIKCATOLVWERMKLLIETAGVGVAAYL 294  
Db 235 PGVLDVNLDEVIVQSSEESIETAKLLALKEGLLVGISSGAAAAAI 281

## RESULT 14

US-09-905-290A-4

; Sequence 4, Application US/09905290A

; Publication No. US20030087327A1

; GENERAL INFORMATION:

; APPLICANT: Rice, John

; APPLICANT: Lanning, Beth

; APPLICANT: Crawford, John  
; APPLICANT: Nye, Gordon  
; TITLE OF INVENTION: METHODS FOR MEASURING CYSTEINE AND DETERMINING CYSTEINE SYNTHA  
; FILE REFERENCE: Docket No. US20030087327A12090US  
; CURRENT APPLICATION NUMBER: US/09/905,290A  
; CURRENT FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; US-09-905-290A-4

Query Match 8.5%; Score 147; DB 9; Length 324;

Best Local Similarity 22.3%; Pred. No. 1.2e-05;

Matches 77; Conservative 62; Mismatches 137; Indels 70; Gaps 13;

QY 12 VEKAHINIRDS----IHLTP-VLTSSILNQLTGRNFFKCELFQKTSFKIRGALNAVR 66  
Db 1 MEEDRCSIKDDATQLIGNTPMVYLNINIVDCVAR-TAAKLEMEPCSSVKERIAYGMKD 59  
QY 67 LVPDALERPKA-VVTHSSGNHQALTYAAKLEGIPAYIVVPQTAPDCKKLAIQAYGASI 125  
Db 60 AEDKGLITPGKSTLIEATSGNTGIGLAFICAAKGYKVLTPMPSSMSLERKIIILLALGAEV 119  
QY 126 VYCEPS-----DESRENVAKRVTEETEGIMVHPNOEPA-----VIAGO 163  
Db 120 HLTDPKSGVOGIIDKAEICSK-----NPDSIMLEOFKNPSNPOTHYTTTGPETWRDSAGE 175  
QY 164 GTIALEVLNQVPLVDALVVPVGGGMLAGIAITVVKALKPSVKVYAAEPSNADDCYQSKLK 223  
Db 176 -----VDILVAGVGTGTLGSGRGLKKNKDFKVVGVPT-----ESAVI 216  
QY 224 GKLMNLYPPETIADG-VKSSIGLNTWPIIRDLVDDIFTVTDEIKCATOLVWERMKLLI 282  
Db 217 SGKPGTHLIQIGAGLIPDLNDFN-----VLDEVIQVTSVEAIEATKALLKEGLLV 269  
QY 283 EPTAGVGVAALVSOHFQTSVPEVKN-----ICIVLSSGNVDLTSSI 323  
Db 270 GISSGAAAAAIK-----VAKRPNAGKLLIVIFPSSGGERYLSTSL 310

## RESULT 15

US-09-931-457A-66

; Sequence 66, Application US/09931457A

; Patent No. US20020157132A1

; GENERAL INFORMATION:

; APPLICANT: Falco, S. Carl

; APPLICANT: Allen, Stephen M.

; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes

; FILE REFERENCE: BB1116 US CIP

; CURRENT APPLICATION NUMBER: US/09/931,457A

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: 09/424,976

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: 60/065,385

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/049,406

; PRIOR FILING DATE: 1997-06-12

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 66

; LENGTH: 386

; TYPE: PRT

; ORGANISM: Solanum tuberosum

US-09-931-457A-66

Query Match 8.5%; Score 147; DB 9; Length 386;

Best Local Similarity 19.2%; Pred. No. 1.5e-05;

Matches 64; Conservative 60; Mismatches 112; Indels 98; Gaps 10;

QY 8 SFADVEKAHI--NIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 65











TOI VOOGTGSVHE TIONENF D IEVI DVQFA  
-----  
QKELSNF LENGJCFKDJD IADG ZS

A97687  
probable serine/threonine dehydratase, degradative (PA2683) [imported] - Aqrobacteriu

C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: AF97687  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: AF97359; PMID:11743194  
A:Accession: AF97687  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88450.1; PID:g15157951; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C4956  
A:Map position: circular chromosome  
C:Superfamily: threonine dehydratase

Query Match 28.6%; Score 496.5; DB 2; Length 324;  
Best Local Similarity 38.7%; Pred. No. 5.6e-31;  
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;

QY 12 VEKAHINIRDSIHLTPVLTSSTLNLTGNLFKCELFKTSFKIRGALNAVRSILVDPDA 71  
DB 7 IEAARERIGNHVRPLTSPFLDEIAGRKLVKAECLQRTGSFKRGWSAVSGLPADV 66

QY 72 LERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIVYCEPS 131  
DB 67 ---RAKVIAFSSGNHQAOGVLAARLHGIPAVIIMPSPDAPKIKIDNTRAYGAEVVLYDRA 123

QY 132 DESRENKAVRTEETEGIMVHPNQEPNAVIAAGGTIALEVLNQ-----VPLVDALVVPVGG 186  
DB 124 NEDRAIGNRLSSERGLTLIRPYDEPLVIAAGGTAGLEIAEQAGELGIGAEEVL-VPCGG 182

QY 187 GGLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYPPEIADG-VKSSSTG 245  
DB 183 GGLTSGISLADAKARNYKVRTAEFERDDVARSAAKIERNATTSGGICDAIVTPQPG 242

QY 246 LNTWPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSPV 305  
DB 243 NITFPIMAGLCCKGIAVSEEEALRAMVLAFLNKLKVVIEPFGGAVALAAL---FHGKELES 299

QY 306 KNICIVLSGGNVD 318  
DB 300 ETVIAVASGGNVD 312

RESULT 11  
AF2912  
threonine dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AF2912  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AF2912  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AA43716.1; PID:g17741247; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: ilvA  
A:Map position: circular chromosome  
C:Superfamily: threonine dehydratase

Query Match 28.6%; Score 496.5; DB 2; Length 324;

Best Local Similarity 38.7%; Pred. No. 5.6e-31;  
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;

QY 12 VEKAHINIRDSIHLTPVLTSSTLNLTGNLFKCELFKTSFKIRGALNAVRSILVDPDA 71  
DB 7 IEAARERIGNHVRPLTSPFLDEIAGRKLVKAECLQRTGSFKRGWSAVSGLPADV 66

QY 72 LERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIVYCEPS 131  
DB 67 ---RAKVIAFSSGNHQAOGVLAARLHGIPAVIIMPSPDAPKIKIDNTRAYGAEVVLYDRA 123

QY 132 DESRENKAVRTEETEGIMVHPNQEPNAVIAAGGTIALEVLNQ-----VPLVDALVVPVGG 186  
DB 124 NEDRAIGNRLSSERGLTLIRPYDEPLVIAAGGTAGLEIAEQAGELGIGAEEVL-VPCGG 182

QY 187 GGLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYPPEIADG-VKSSSTG 245  
DB 183 GGLTSGISLADAKARNYKVRTAEFERDDVARSAAKIERNATTSGGICDAIVTPQPG 242

QY 246 LNTWPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSPV 305  
DB 243 NITFPIMAGLCCKGIAVSEEEALRAMVLAFLNKLKVVIEPFGGAVALAAL---FHGKELES 299

QY 306 KNICIVLSGGNVD 318  
DB 300 ETVIAVASGGNVD 312

RESULT 12  
H84359  
threonine dehydratase [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: H84359  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leichauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: H84359  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-495 <STO>  
A:Cross-references: GB:AE004437; NID:g10581515; PIDN:AAG20244.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: iluA

Query Match 28.0%; Score 486.5; DB 2; Length 495;  
Best Local Similarity 36.7%; Pred. No. 6e-30;  
Matches 116; Conservative 59; Mismatches 130; Indels 11; Gaps 7;

QY 7 ISFADVERAHINIRDSIHLTPVLTSSTLNLTGNLFKCELFKTSFKIRGALNAVRS 66  
DB 87 LSLSDIARDARVSEVTSHTFSDLTGADVRLKLECFORTGTSFKIRGATNRI 146

QY 67 LVPDALKERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIAQYASIV 126  
DB 147 LSAD---QDAGVVTASAGNHAQGVALAASRGVSKVVMPEASAPISKIKATKSYGAEVV 203

QY 127 YCEPDESRENKAVRTEETEG-IMVHPNOEPAVIAAGGTIALEVLNQVPLVDALVVPV 185  
DB 204 -LHGADYDDAOAHAELEAAEGRTVYHAFDVEYIMAGGTGLGLEIAADCTVDTVVVPIG 262

QY 186 GGGMLAGIAITVKALKPSVKVYAAEPSNADDCVQSKLKGKLMPLNLYP-PETTADGVK-SS 243  
DB 263 GGGLSGVATALKGELDDVRVIGVQAEGASTVARSLLDKG---POAVDHVDITDIADGIVARQ 320

QY 244 IGLNTPPIIRDVLDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVGAVALSQHFQTVSP 303  
DB 321 VQAQTFPVIOEHVDEVTVSDDIEATLVLLERGLTVEGAGATATAAVALDEKFEYADG 380

QY 304 EVKNICIVLGGNVDL 319  
 Db 381 E--TIVPALCGGNIDL 394

RESULT 13  
 D90166  
 hypothetical protein tdcB [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Jun-2002  
 C:Accession: D90166  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: D90166  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-405 <KUR>  
 A:Cross-references: GB:AE006641; NID:gl13813385; PIDN:AAK40587.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: tdcB  
 C:Superfamily: threonine dehydratase

Query Match 27.8%; Score 483; DB 2; Length 405;  
 Best Local Similarity 38.7%; Pred. No. 8.5e-30;  
 Matches 128; Conservative 62; Mismatches 115; Indels 26; Gaps 11;

QY 5 YCISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAV 64  
 Db 3 YLEFDRIRLAKEIKYVHIPIIDYSTFSRIINAKVYLKLENLQKTSFKVRGAFNKL 62

QY 65 RSLVPDALEKPKAVVTHSSGNHQALTYAAKLEGIPAYIVVPOAPDCKKLAIAQYAS 124  
 Db 63 LSLKE---BEKNGVIASAGNHAQGVAAATLNIKSTIVMPETAPASKYLATRSYGA 119

QY 125 IV-YCEPSDESRENKVRTEETEGIMVHPNPEAVIAGOGTIALEVLNPLVDALVVP 183  
 Db 120 VVLYGKYLHESKK-AEELIQNTGLFVHPYSDLDVITGGGIGLELDIEP-DYVIIP 176

QY 184 VGGGMLAGIAITVKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNPPETIADG--VK 241  
 Db 177 IGGGGLGISIALKYRPNVKIIGVQSSSPSMKVSXDLGRV-EIEPSYIADGILVK 235

QY 242 SSIGLNTWPIIRDVDDIFTTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQTV 301  
 Db 236 SPSEL-TFSIINELVDDIVLVDDEIAEAIVLLERSKTLAE---GAGAAALAS----LI 287

QY 302 SPEV-----KNICIVLGGNVDLT--SSIT 324  
 Db 288 SKGVKVGNDKVKVLSVSGGNIDLSLLTLT 318

RESULT 14  
 G87633  
 threonine dehydratase [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 31-Dec-2001  
 C:Accession: G87633  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon-  
 n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete genome sequence of Caulobacter crescentus.  
 A:Reference number: AB8749; MUID:21173698; PMID:11259647  
 A:Accession: G87633  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-325 <STO>  
 A:Cross-references: GB:AE005673; NID:gl13424763; PIDN:AAK25067.1; GSPDB:GN00148  
 C:Genetics:

A:Gene: CC3105  
 C:Superfamily: threonine dehydratase

Query Match 27.6%; Score 479.5; DB 2; Length 325;  
 Best Local Similarity 34.7%; Pred. No. 1.2e-29;  
 Matches 110; Conservative 62; Mismatches 134; Indels 11; Gaps 4;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRS 66  
 Db 3 VSLADIQAAAVRLKSAVETPLIESPALNDRUGGRIFLKPETLQAGAFKFRGAYNRUSQ 62

QY 67 LVPDALEKPKAVVTHSSGNHQALTYAAKLEGIPAYIVVPOAPDCKKLAIAQYASIV 126  
 Db 63 L---SDEEKARGVAFSSGNHAQGVAAARLLGVLPALVWPSDSVSKVEGTGFGADIR 119

QY 127 YCEPSDESRENKVRTEETEGIMVHPNPEAVIAGOGTIALEVLNPLVDALVVP 182  
 Db 120 FYDRETFEDRVATADQIAAERGCVVPSYDDPHIIAGOGTVGLEIVAAQAAOGATLDRLIC 179

QY 183 PVGGGMLAGIAITVKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNPPETIADGVKS 242  
 Db 180 CVGGGGLIAGTSTAVKALSPATEIWMGVEPAGEDETRRSLESGRRRETIDKARSICDALIT 239

QY 243 SI-GLNTWPIIRDVDDIFTTVEDEIKCATQLVWERMKLLIEPTAGVGVAAVLSQHFQTV 301  
 Db 240 PIPGDLTPINQKNSLGGVVAVTDAEVAEMRYAFSTLKLIVPEPGCVALTALTCKVDVA 299

QY 302 SPEVKNICIVLGGNVD 318  
 Db 300 G---KTVAILVLSGGNVD 313

RESULT 15  
 AB0897  
 threonine amonia-lyase (EC 4.3.1.19) [similarity] - Salmonella enterica subsp. enter  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 02-Aug-2002  
 C:Accession: AB0897  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AB0897  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-329 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD07768.1; PID:gl6504317; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY3427  
 C:Superfamily: threonine dehydratase  
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase

Query Match 27.1%; Score 471; DB 2; Length 329;  
 Best Local Similarity 36.6%; Pred. No. 5.4e-29;  
 Matches 123; Conservative 64; Mismatches 121; Indels 28; Gaps 10;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRS 66  
 Db 9 VAIEDILEAKKRLACKIYKVTGMPRSNYFSERCKGEIFLKFENMORTGSEFKIRGAFNKLSS 68

QY 67 LVPDALEKPKAVVTHSSGNHQALTYAAKLEGIPAYIVVPOAPDCKKLAIAQYASIV 126  
 Db 69 LTE---AEKRGVAVACSTGNHAQGVSLSCAMLGIDGKVMKPKGAPSKVAATCDYSAEVV 125

QY 127 YCEPSDESRENKVRTE--ETEG-IMVHPNPEAVIAGOGTIALEVLNPLVDALVVP 183  
 Db 126 L--HGDNFNDIAK-VSEIVETEGRIFIPDYDDPKRVAGGGTIGLEIMEDLDYDNDVIVP 182

QY 184 VGGGMLAGIAITVKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNPPETIADGVK--S 242

Search completed: June 24, 2003, 06:23:10  
Job time : 28.0398 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 05:44:58 ; Search time 41.5611 Seconds  
(without alignments)  
1685.614 Million cell updates/sec

Title: US-09-889-609B-10  
Perfect score: 1735  
Sequence: 1 MCAQYCISFADVEKAHINR.....SSITWQKAERPASYQSVSV 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	802.5	46.3	346	10 Q9T0D1	Q9T0D1 arabidopsis
2	602	34.7	323	16 Q8XPX0	Q8XPX0 ralstonia s
3	550	31.7	388	3 Q8X0J0	Q8X0J0 neurospora
4	539	31.1	320	16 Q910F5	Q910F5 pseudomonas
5	535	30.8	325	16 Q9XAA4	Q9XAA4 streptomyces
6	532.5	30.7	401	16 Q9WVJ1	Q9WVJ1 thermotoga
7	517	29.8	346	16 Q9U500	Q9U500 staphylococ
8	514.5	29.7	317	5 Q93968	Q93968 caenorhabdi
9	508.5	29.3	323	16 Q921F8	Q921F8 rhizobium m
10	499	28.8	402	16 Q8XL77	Q8XL77 clostridium
11	497.5	28.7	404	17 Q975Y3	Q975Y3 sulfolobus
12	496.5	28.6	324	16 Q8UBW5	Q8UBW5 agrobacteri
13	486.5	28.0	409	16 Q9EWG3	Q9EWG3 streptomyces
14	486.5	28.0	495	17 Q9NNH6	Q9NNH6 halobacteri
15	483	27.8	405	17 Q980P1	Q980P1 sulfolobus
16	480.5	27.7	406	17 Q97CB2	Q97CB2 thermoplasm

Q9A3U7 caulobacter  
Q9HLW2 thermoplasm  
Q8RDT9 fusobacteri  
Q8RVF0 pyrobaculum  
Q8Z3K4 salmonella  
Q92HZ2 rickettsia  
Q916G0 pseudomonas  
Q8Z9V3 yersinia pe  
Q9VHF0 drosophila  
Q987A3 rhizobium l  
Q9PP95 campylobact  
Q8UP8 agrobacteri  
Q9ZD93 rickettsia  
Q9RWU8 deinococcus  
Q21080 caenorhabdi  
Q95XY8 caenorhabdi  
Q92WC4 rhizobium m  
Q94634 schizosacch  
P73375 synecocyst  
Q9A2D0 caulobacter  
Q9YBV1 aeropyrum p  
Q985M4 rhizobium l  
Q98Q4 rhizobium l  
Q91418 pseudomonas  
Q9AUQ1 oryza sativ  
Q8W314 oryza sativ  
Q9XBI2 bacillus ce  
Q8ZAB4 yersinia pe  
Q9KWI1 vibrio chol

#### ALIGNMENTS

#### RESULT 1

Q9T0D1 ID Q9T0D1 PRELIMINARY; PRT; 346 AA.  
AC Q9T0D1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE TSC23.70 protein.  
GN TSC23.70 OR AT4G11640.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL049500; CAB39935.1;  
DR EMBL; AL161532; CAB78207.1;  
DR HSP; P04968; ITDJ  
DR InterPro; IPR001926; B6\_enzyme\_beta.  
DR InterPro; IPR000634; S/T\_dehydrtse.  
DR Pfam; PF00291; PALP; 1.  
DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.

```

KW Hypothetical protein.
SQ SEQUENCE 346 AA; 36839 MW; 47C37E1CA52B91F0 CRC64;

Query Match
Best Local Similarity 46.3%; Score 802.5; DB 10; Length 346;
Matches 163; Conservative 67; Mismatches 83; Indels 23; Gaps 6;

QY 4 QYCISFADVEKAHINIRDSIHLTPVLTSILNLTGRNLFKCELFQ----- 50
DB 7 KYAADILSIKEAHRDIPIYIHTPTVLTSELSISGRSLFFKCECLQKIGIECRSFYSVY 66
QY 51 --KTGSFKIRGALNAVRSLVPDALKRPKAVVTHSSGNHGQALTYAAKLEGIPAYIVPQ 108
DB 67 CKSGAFKRGACNAVLSL--DA-EQAGKGVVTHSSGNHAAALSLAAKIQIGIPAYIVPQ 123
QY 109 TAPDCKKLAIAQYASIVYCEPSDESRENKAVRTEETEGIMVHPNOEPAVIAQGGTIAL 168
DB 124 GAPCKVDNVIRYGGKVIWSEATSSREIAASKVLOETGSLIHPYNDGRIISQGGTIAL 183
QY 169 EVLNQVPLVDALVVPVGGGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLM 228
DB 184 ELLEQIQIIDAIVVPISGGLISGVALAAKSIKPSIRIIAAEPKGAADAAQSKVAGKII- 242
QY 229 NLYPPTIADGVKSSIGLNTWPIIRDLVDITFTTEDEIKATOLVWERMKLLIEPTAGV 288
DB 243 TLPVTNTIADGLRASLDGLTPWVVRDLVDVITTECEITEAMKMCYEILKVSVEPSGAI 302
QY 289 GVAAVLSGHFTVSP---EVKNICIVLSGGNVDLTS 321
DB 303 GLAAVLSNFRN-NPSCRDCKNIGIVLSGGNVDLGS 337

RESULT 2
Q8XPX0 PRELIMINARY; PRT; 323 AA.
AC Q8XPX0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Probable amino-acid dehydratase protein (EC 4.2.-.-).
GN RSP1516 OR RS04793.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18667.1;
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
DR Lyase; Plasmid; Complete proteome.
SQ SEQUENCE 323 AA; 34515 MW; 2175CBF693E04A2D CRC64;

Query Match
Best Local Similarity 34.7%; Score 602; DB 16; Length 323;
Matches 128; Conservative 65; Mismatches 113; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINIRDSIHLTPVLTSILNLTGRNLFKCELFQKTSFKIRGALNAVR 66
DB 6 ITVDVVAHAARLQGVHRTPTVLTSTTANALTGAELFFKCNCFQMGAFKRGYNALSG 65
QY 67 LVPDALERKPAVVTHSSGNHGQALTYAAKLEGIPAYIVVPTAPDCKKLAIAQYASIV 126

Db 66 FTP-----EQRKAGVITFSSGNHQAIALSARLLGMRAVIVMPKDAPIKVEATRGYGGVW 122
QY 127 YCESPDESRENKAVRTEETEGIMVHPNOEPAVIAQGGTIALVNLQVPLVDALVVPVGG 186
DB 123 FYDRTEDREAIGRRRLAEQHGTLTLPYDHPVWAGOGCTAAKELIEETGPDLDLLVCGG 182
QY 187 GGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMNLYPPETIADGVKSS-IG 245
DB 183 GGLLSGCGATAARALSPACRIIGVEPEAGNDGQOQSLRKEIV-HIDTPATIADGAQTQHLG 241
QY 246 LNTWPIIRDLVDITFTTEDEIKATOLVWERMKLLIEPTAGVGVAAVLSQHFQTVSPV 305
DB 242 HYTEAVIRALVDDIATVSDADLVDMRFFAGRMKIVVEPTGCLAAALRQR---VEVRG 298
QY 306 KNICIVLSGGNVDL 319
DB 299 KRGVIIISGGNVDL 312

RESULT 3
Q8X0J0 PRELIMINARY; PRT; 388 AA.
AC Q8X0J0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Related to threonine dehydratase.
GN BLH24.140.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schultze U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670005; CAD21283.1;
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
SQ SEQUENCE 388 AA; 41505 MW; 02BFB864B95716A6 CRC64;

Query Match
Best Local Similarity 31.7%; Score 550; DB 3; Length 388;
Matches 140; Conservative 62; Mismatches 109; Indels 58; Gaps 10;

QY 10 ADVEKAHINIRDSIHLTPVLTSILNLTGR-----NLEFKC 46
DB 13 ASVLAASHLKPYIHTPTVLTNTTTFDOLASTPRTPEELQGTWEATERPANPKIRFWFKC 72
QY 47 ELFKTGSFKIRGALNAVRSL-VPDALE-RKPAVTHSSGNHGQALTYAAKLEGIPAYI 104
DB 73 ENQIRIGAFKARGAFHAFVERLKOTEGLEGKGVVTHSSGNHQAALSLAARENGI PAHI 132
QY 105 VVPOTAPDCKKLAIAQYASIVYCEPSDESRENKAVRTEETEGIMVHPNOEPAVIAQGG 164
DB 133 VMPTISPPPKIAATKGYGANITFSGSTSTEREAVTREVIEKTGARLVPPYDHPDILQGG 192
QY 165 TIALEVLNVOPL-----VDALVVPVGGGMLAGTATVTKALPKS----- 203
DB 193 TAAELQRQAASLSGSTGTTTNRRLNIAITPCGGGGLLSGTALACSLDSPSPSTPG 252
QY 204 -KVYAAEP---SNADDCYQSKLKGKLMNLYPPETIADGVKSSIGLNTWPII--RDLVDD 258
DB 253 PILVFGAEPSPSGADGRRGYSGSTRIESV-KSLTADGLRTPLGAYPNSIIYERKLVAG 311
QY 259 IFTVTEDEIKATOLVWERMKLLIEPTAGVGVAAVLSQHFQTV-----SPEVKNICIVL 312

```

```

Db 312 MYSGVEEIKKALRLVYERMKVYVPEPSAVVGLVAFNFEFRSMVEREGGEGWDLGVVF 371
QY 313 SGGNVDLTS 321
Db 372 SGGNVELAA 380

RESULT 4
Q910F5
ID Q910F5 PRELIMINARY; PRT; 320 AA.
AC Q910F5;
DT 01-WAR-2001 (TREMBLrel. 16, Created)
DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-WAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable serine/threonine dehydratase, degradative.
GN PA2683.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Miziouchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saiter M.H., Hancock R.E.W., Iori S., Olson M.V.;
RA "complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004696; AAG06071.1; -
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
KW Complete proteome.
SQ SEQUENCE 320 AA; 33957 MW; C6F97C6259FE8584 CRC64;

Query Match 31.1%; Score 539; DB 16; Length 320;
Best Local Similarity 38.0%; Pred. No. 4e-35;
Matches 119; Conservative 66; Mismatches 120; Indels 8; Gaps 4;

QY 8 SPADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVRSL 67
Db 6 TYDDVIAAARTAGHANRTPVMSRSLDEELGAEEVFFKCNLQRMGAFFRGAFNALSRF 65
QY 68 VPDALERKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTAPDCKKLAIQAYGASIVY 127
Db 66 ---SAEORAAAGVAFSSGNAQAIALSARLLGIPATVMPADAPAVKTEATRGYGQVVL 122
QY 128 CEPSPDESRENKRVTEETEGIMVHPNCEPAVIAQCGTIALEVLNQVPLVDALVVPVGG 187
Db 123 YDRYTEDREQIGRDIAQRHGLTLIPPYDHPDLVAGQGAELFEEVGLPDFAFFPLGGG 182
QY 188 GMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPETIADGVKSS-IGL 246
Db 183 GLLSCALAIRALAPACRIYGEPEAGNDGQSRSLSGAIV-HIDTPQTLADGAQTOHLGN 241
QY 247 NTPWIIRDLVDDIFTVTDEIKCATQLVWERMKLLIETAGVGVAAVLSQHPQTYSPEVK 306
Db 242 LTFPLIQRNVDDLITASDAELVDGMRFLAARMKLLVEPTGCLGLAAARQKDEL---RGK 298
QY 307 NICIVLSGGNVDL 319
Db 299 RVGILLSGGNVDL 311

RESULT 5
Q9XAA4
ID Q9XAA4 PRELIMINARY; PRT; 325 AA.
AC Q9XAA4;

```

```

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE Putative threonine dehydratase.
GN SC00821 OR SCF43A.11C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL096837; CAB48898.1; -
DR HSSP; P04968; 1TDJ.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 325 AA; 33965 MW; 4E1DE62D340489A7 CRC64;

Query Match 30.8%; Score 535; DB 16; Length 325;
Best Local Similarity 39.6%; Pred. No. 8.5e-35;
Matches 124; Conservative 56; Mismatches 125; Indels 8; Gaps 4;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNQLTGRNLFKCELFQKTSFKIRGALNAVR 66
Db 8 VTLDVRSAAARIKGVANRTPVLRSTLDALVGAEVHLKCENQORVGAFFRGAYNAASR 67
QY 67 LVPDALERKPKAVVTHSSNGHQALTYAAKLEGIPAYIVVPOTAPDCKKLAIQAYGASIV 126
Db 68 LTPQELAR---GTAAYSSGNAQAVALARELGTAVIVMPEDAPPSPKRDATAGYAEIV 124
QY 127 YCEPSPDESRENKRVTEETEGIMVHPNCEPAVIAQCGTIALEVLNQVPLVDALVVPVGG 186
Db 125 TYDRYTGDRVAVAEALADRGTLTLPPEYEHPIVAGQGAELFEEVGLPDFAFFPLGGG 184
QY 187 GMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPETIADGVKSS-IGL 245
Db 185 GGLIAGSATAVKALHPGHRVIGVEPEAGDDTTRSLAAGRRV-SVVPRTIADGQALPTPG 243
QY 246 LNTWPIIRDLVDDIFTVTDEIKCATQLVWERMKLLIETAGVGVAAVLSQHPQTYSPEV 305
Db 127 YCEPSPDESRENKRVTEETEGIMVHPNCEPAVIAQCGTIALEVLNQVPLVDALVVPVGG 186

```



RC	STRAIN=1021;	RC	MEDLINE=21396509; PubMed=11481432;
RX	Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,	RA	Barlow-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA	Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,	RA	Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA	Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;	RT	"Nucleotide sequence and predicted functions of the entire
RT	<i>Sinorhizobium meliloti</i> pSYMA megaplasmid."	RL	Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
RL	EMBL; AE007289; AAK65688.1; ..	DR	InterPro: IPR001926; B6_enzyme_beta.
DR	InterPro: IPR000634; S/T_dehydrtse.	DR	PFam; PF00291; PALP; 1.
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN.1.	DR	Plasmid; Hypothetical protein; Complete proteome.
KW	SEQUENCE 323 AA; 34706 MW; 1BF363B67F4D635B CRC64;	DR	
QY	Query Match 29.3%; Score 508.5; DB 16; Length 323;	DR	Best Local Similarity 36.8%; Pred. No. 1.1e-32;
DB	Matches 114; Conservative 62; Mismatches 127; Indels 7; Gaps 4	DR	
QY	11 DVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFKFCELFQKTSFKIRGALNAVRSLLVPD 70	DB	9 DIREARLRPHVHRHTPLLRAEKTEKAAGCOLYLPETLQITGAFKIRGALKALS---- 65
QY	71 ALERKPKAVYTHSSGNHGOALTYAAKLEGIPAYIVVPTAPDCKKLAIOYGASIVYCEP 130	DB	66 SREIANGIIATSSGNAAGLSYAAKMLGVKVLVLPVTTTPKIKENTKALGAEVILFDG 125
QY	131 SDSESRNAVKKRVTEETEGIMVHPNQEPAVIAGQGTIALEVLNQVLVDALVVPVGGGML 190	DB	126 DNAARNKKVYEIABGNKYAVIHGFEDPVMVMAQGTIGCEILDDVDVTIVPLGGGGLI 185
QY	191 AGTAITVWALKPSVKVYAAEPSNADDCYQSKLKGKMLPNLYPPETIADGVKSSI-GLNTW 249	DB	186 SGIATAIKETKPSVRVIGAEPAITPKYFHSRV-NKERTSLPLKNTIADGLRISVPCQNPY 244
QY	250 PIIRDLVDDLTFTVTEDEIKATOLVWERMKLLIEPTAGVGAVALLSQHFQTVSPVKNIC 309	DB	245 PIIEKYVDEIVLVEDEHIIAGMRALAKDAKUIAEPAASIGVALLAGIIDVKLDE--KVC 302
QY	310 IVLSSGGNVLD 319	DB	303 AVLSSGGNWDL 312
RESULT 10		DB	
Q8XL77	PRELIMINARY; PRT; 402 AA.	DB	
ID	Q8XL77	DB	
AC	Q8XL77	DB	
DT	01-MAR-2002 (TREMBlrel. 20, Created)	DB	
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)	DB	
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)	DB	
DE	Threonine dehydratase.	DB	
GN	THD OR CPE1165.	DB	
OS	Clostridium perfringens.	DB	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;	DB	
OC	Clostridiales; Clostridiaceae; Clostridium.	DB	
ON	NCBI_TaxID=1502;	DB	
OX	[1]	DB	
RP	SEQUENCE FROM N.A.	DB	
RC	STRAIN=13 / TYPE A;	DB	
RC	pubMed=11792842;	DB	
RA	Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,	DB	
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;	DB	
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic	DB	
RT	flesh-eater";	DB	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).	DB	
DR	EMBL; AP003189; BAB80871.1; ..	DB	
DR	InterPro: IPR002912; ACT.	DB	
DR	InterPro: IPR001926; B6_enzyme_beta.	DB	
DR	InterPro: IPR000634; S/T_dehydrtse.	DB	



```
Matches 121; Conservative 48; Mismatches 131; Indels 13; Gaps 5;
QY 12 VEKAHINIRDSIHLPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRSVLPDA 71
Db 7 IEAARERIGHAVRTPLLTSPFLDETAGRKLKFAECLORTGSFKFRGWSAVSGLPADV 66
QY 72 LERPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPDCCKKLAIQAYGASIVYCEPS 131
Db 67 ---RAKGVIATFSGNHAQGVVALAARLHGIPAVIIMPSPDKIKIDNTRAYGAEBVLYDRA 123
QY 132 DESRENVAKRVTEETEGIMVHPNOEPVAVIAGQGTIALEVLNQ-----VPLVDALVVPVG 186
Db 124 NEDRDAGNRKLSERGLTIRPYDEPLVIAGQGTAGLETAEOGAELGIGAAEVL-VPCGG 182
QY 187 GGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADG-VKSSIG 245
Db 183 GGLTSGISLALDAKARNYKRVTAEPERFDDVARSAAAGKIERNATVSGSICDAIVTPQG 242
QY 246 LNTWPIIRDLDVDDITFTVTEDEIKCATOLVWERMKLLIETPTAGVGVAIVLSQHFTQVSPEV 305
Db 243 NITFPIMAGLCGRGIAVSEELRAMVLAFNRLKVVIEPGAVALLAAL---FHGKELES 299
QY 306 KNICIVLSGGNVD 318
Db 300 ETVIAVSGGNVD 312

RESULT 13
Q9EWG3 PRELIMINARY; PRT; 409 AA.
AC Q9EWG3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative amino acid deaminase.
GN 25CK31.22 OR SCO4962.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Harris D.E., Ouail M.A., Kieser H.,
RA Thomson N.R., James K.D., Chatter K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
```

```
RL Nature 417:141-147(2002).
DR EMBL; AL451182; CAD18697.1; -
DR AL512667; CAD30948.1; -
DR HSSP; P04968; 1TDJ
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse.
DR Pfam; PF00291; PALP; 1.
DR TIGRFAMS; TIGR01127; ilva.1cterm; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 409 AA; 42279 MW; 9F21A40D3E9A9B9 CRC64;

Query Match 28.0%; Score 486.5; DB 16; Length 409;
Best Local Similarity 38.2%; Pred. No. 9.1e-31;
Matches 120; Conservative 54; Mismatches 129; Indels 11; Gaps 6;

QY 7 ISFADVEKAHINIRDSIHLPVLTSSILNLTGRNLFKCELFQKTSFKIRGALNAVRS 66
Db 12 VTLDVDRGAKMLSGVARVTAMEGRHLSQVSPVHLKCKENLQRTGSFKLRGAYVRIAG 71
QY 67 LVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPOTAPDCCKKLAIQAYGASI - 125
Db 72 LLP---EQRAGVVAASAGNHAQGVVALAARLHGIPAVIIMPSPDKIKIDNTRAYGAEBV 128
QY 126 VYCEPSDESRENVAKRVTEETEGIMVHPNOEPVAVIAGQGTIALEVLNQVPLVDALVVPVG 185
Db 129 LHGVVVDETLA-AAQEVAERTGAVFIHPDPHDVIAQGQVGLGLEIQCEPEVTVVVGIG 187
QY 186 GGMLAGIATVTKALPKSVKYAAEPSNADDCYQSKLKGKLMPLNLYPPETIADGVK-SSI 244
Db 188 GGGLAAGIAVAVKALRPDVRIVGVQAEGA-AAYPSPSLAAGRPAVAVONPATMADGIKVG 246
QY 245 GLNTWPIIRDLDVDDITFTVTEDEIKCATOLVWERMKLLIETPTAGVGVAIVLSQHFTQVSPE 304
Db 247 GDVPFGIVGELVDEVTVSSEDALSTALLCLERAKLVVEPAGASPVAAALLSEPDDEGP - 305
QY 305 VKNICIVLSGGNVD 318
Db 306 ---VWAVLSGGNVD 316

RESULT 14
Q9HNH6 PRELIMINARY; PRT; 495 AA.
AC Q9HNH6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Threonine dehydratase.
GN ILUA OR VNG2100G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005099; AAG20244.1; -
DR HSSP; P04968; 1TDJ
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001926; B6_enzyme_beta.
DR Pfam; PF000634; S/T_dehydrtse.
DR Pfam; PF01842; ACT; 1.
```

